Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation

Gangqing Hu1,3, Qingsong Tang1,3, Suveena Sharma2,3, Fang Yu2, Thelma M Escobar2, Stefan A Muljo2, Jinfang Zhu2 & Keji Zhao1

Although intergenic long noncoding RNAs (lincRNAs) have been linked to gene regulation in various tissues, little is known about lincRNA transcriptomes in the T cell lineages. Here we identified 1,524 lincRNA clusters in 42 T cell samples, from early T cell progenitors to terminally differentiated helper T cell subsets. Our analysis revealed highly dynamic and cell-specific expression patterns for lincRNAs during T cell differentiation. These lincRNAs were located in genomic regions enriched for genes that encode proteins with immunoregulatory functions. Many were bound and regulated by the key transcription factors T-bet, GATA-3, STAT4 and STAT6. We found that the lincRNA LincR-Ccr2-5′AS, together with GATA-3, was an essential component of a regulatory circuit in gene expression specific to the Th2 subset of helper T cells and was important for the migration of Th2 cells.

The mammalian genomes encode tens of thousands of long noncoding RNAs (lincRNAs)1,2. These transcripts have essential roles in regulating gene expression and affect various biological processes during development and in pathological conditions3,4. One classic example of a functional intergenic IncRNA (lincRNA) is encoded by Xist, located on the X chromosome; it is required for inactivation of the X chromosome in females, and it operates by recruiting repressive complexes such as PRC2 to the silenced X chromosome5. Another well-characterized example is HOTAIR, which recruits PRC2 complexes to homeobox (Hox) domains and represses expression of HOXD (which encodes HOX4)6. Additionally, several other lincRNAs function to mediate methylation of histone H3 at Lys27 by recruiting PRC2 (refs. 7,8). Studies have also reported IncRNAs with enhancer functions9. LincRNAs may also regulate gene expression through post-transcriptional mechanisms10,11.

The study of lincRNA function in the immune system is an emerging field. Helper T cells are critical for orchestrating adaptive immune responses to a variety of pathogens; they are also involved in the pathogenesis of various types of immunological diseases, including allergy, asthma and autoimmunity12. The lincRNA TMEVPG1 (NeST; called ‘LincR-Ifng-3′AS’ here) is reported to be specifically expressed by the Th1 subset of helper T cells and is critical for controlling infection with Th11, which encodes interferon-γ (encoded by Ifng)14. TMEVPG1 interacts with WDR5, a core subunit of the MLL histone H3 Lys4 (H3K4) methyltransferases, and facilitates histone methylation at the Ifng locus in CD8+ T cells15. A survey of IncRNA in CD8+ T cell from mouse spleen by custom array has suggested a pivotal role for IncRNA in the differentiation and activation of lymphocytes16.

Despite those examples, the function and transcriptional regulation of lincRNAs during the development and differentiation of T cells is far from understood, partially because of the lack of knowledge about lincRNA expression in cells of the immune system17. Thus, to better understand the role of lincRNAs in the development and differentiation of T cell lineages, we applied high-throughput sequencing technologies to sequence cDNA (RNA-Seq) of 42 subsets of thymocytes and mature peripheral T cells at multiple time points during their differentiation. Analysis of this data set identified 1,524 genomic regions that generate lincRNAs. Our data revealed a highly dynamic and cell- or stage-specific pattern of lincRNA expression. Genomic location analysis of the genes encoding lincRNA revealed that they were adjacent to genes encoding proteins critically involved in regulating immunological function, which suggested possible coevolution of protein-encoding and lincRNA-encoding genes. Through the use of gene-deficient mice, we found that the transcription factors T-bet, GATA-3, STAT4 and STAT6 accounted for the cell-specific expression of most lincRNAs in Th1 and Th2 cells. Inhibition of a Th2-specific lincRNA we have called ‘LincR-Ccr2-5′AS’ (nomenclature discussed below), whose expression is regulated by GATA-3, by short hairpin RNA (shRNA) resulted in the deregulation of many genes ‘preferentially’ expressed in Th2 cells, including several chemokine receptor–encoding genes located in the vicinity of the gene encoding LincR-Ccr2-5′AS, and compromised the migration of Th2 cells to

1Systems Biology Center, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA. 2Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA. 3These authors contributed equally to this work. Correspondence should be addressed to J.Z. (jfzhu@niaid.nih.gov) or K.Z. (zhaok@nhlbi.nih.gov).
lung tissue in mice. Therefore, our data sets provide a comprehensive resource for future studies of the function and mechanisms of lincRNA in the development, differentiation and immune response of T cells.

RESULTS
Cataloging 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), CD4+CD8+ double-positive (DP) and CD4+ or CD8+ single-positive (SP) thymic T cells, and thymus-derived regulatory T cells (iTreg cells) from the lymph nodes of C57BL/6 mice. Additionally, we obtained T11, T12, T17 and induced Treg cells (iTreg cells) through in vitro differentiation of naive CD4+ T cells for a various length of time in culture. In total, we obtained 42 T cell subsets (each with biological duplicates; Supplementary Fig. 1a). We analyzed total and/or polyadenylated (poly(A)) RNA from those cells by RNA-Seq. Using a strategy similar to one already published8 (Supplementary Fig. 1b), we identified a total of 1,524 lincRNA-expressing genomic regions (or clusters) in the 42 T cell subsets (Supplementary 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 Gata3 contained at least two divergently transcribed lincRNA-encoding genes (Fig. 1a). We were unable to identify 73% of the clusters by noncoding gene annotations from public databases such as RefSeq19, Ensembl20, UCSC21 and NONCODE22 and thus they were previously unknown potential lincRNA-encoding genes (Supplementary Table 1). The number of lincRNA clusters identified in each T cell subset ranged from 154 to 354 (Fig. 1b).

We used several criteria to estimate the coding potential of those potential lincRNA-encoding gene clusters. We used transcriptome assembly to retrieve transcripts (both spliced and unspliced) from each cluster. Approximately 482 clusters (32% of all clusters) encoded spliced transcripts. On average, each lincRNA-encoding gene cluster encoded 1.5 independent transcripts (spliced or unspliced but originating from the same promoter); there were a total of 2,194 such transcripts. Comparison of the putative open reading frames of those transcripts against the Swiss-Prot protein database with the basic local alignment search tool revealed that only 72 clusters encoded transcripts with significant similarity to protein-encoding genes (E value < 10−5; identity > 30%). We also evaluated the coding potential of those 2,194 transcripts by using the Coding Potential Calculator23. This detected transcripts with high coding potential in 91 clusters.

Figure 1 Identification and lineage-specific expression of lincRNAs. (a) Position of the LincR-Gata3-3′ cluster on chromosome 2 (Chr 2) containing two lincRNAs from the Watson and Crick strands, assessed by strand-specific RNA-Seq; blue upward arrowheads, promoters; vertical axis, number of ‘reads’ divided by total ‘read’ number and by window size, presented as ‘reads’ per genomic position per million ‘reads’ (RPMS). (b) Quantification of lincRNAs in DN cells (stages DN1-DN4), DP cells (stages DP1 and DP2), CD8+ SP T cells, iTreg T cells and CD4+ T cells harvested ex vivo (CD4+, thymic CD4+ T cells; Naive CD4+, naive CD4+ T cells from lymph nodes (Supplementary Fig. 1a)), and subsets of T11, T12, T17 and iTreg cells obtained in vitro after 2 weeks of polarization; numbers above bars indicate total for each subset. (c,d) Heat maps of lincRNAs and mRNAs with a difference in expression of over twofold (false-discovery rate, <0.01) in any two subsets of T cells (c) or in naive CD4+ T cells versus distinct helper T cells at 2 weeks of polarization (d). (e–g) Cell-specific and common lincRNAs (top row) and mRNAs (bottom row) among various DN cells (e), DP, SP and iTreg cells (f), and helper T cell subsets (g), assessed by total RNA-Seq and presented as Venn diagrams; numbers around perimeter indicate frequency of cell-specific genes (purple) and common genes (blue). (h) Hierarchical cluster analysis of lincRNA expression in naive CD4+ T cells and T11, T12, T17 and iTreg cell subsets after 2 weeks of polarization, assessed by total RNA-Seq: each column represents one lincRNA cluster, whose expression values were transformed into z-scores; ‘X-pref’ indicates a lincRNA with expression in lineage X 1.5-fold higher than the maximum in other lineages; ‘shared’ indicates all other conditions. Data are representative of two experiments (biological replicates).
Finally, transcripts from lincRNA-encoding genes expressed in T\(_{\text{H}2}\) cells (at 2 weeks of polarization) showed more significant enrichment in the nucleus than in the cytoplasm compared to transcripts from protein-encoding genes (Supplementary Fig. 1c), which suggested that most of the lincRNAs were not translated. These analyses indicated that the majority of the lincRNA-encoding loci identified (>90%) had limited coding potential. Because over half of the clusters encoding potential coding transcripts also encoded noncoding transcripts (data not shown), we included them for further data analysis.

To provide a systematic identifier for each lincRNA, we proposed the following nomenclature: LincR-‘name’ or LincR-‘name‘ for a lincRNA cluster situated near protein-encoding gene ‘name’. The additional designation ‘S’ (sense) or ‘AS’ (antisense) may be added if the direction of the lincRNA cluster relative to that of gene X can be inferred from the transcript assembly. Where more than one lincRNA could have the same name, their distance (in kilobases) from gene X is further appended (such as ‘LincR-Chd2-5′-74K’, where ‘74K’ indicates it is 74 kilobases from Chd2). In summary, by analyzing data sets of T cells at various stages and differentiation stages through RNA-Seq and chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq), we identified 1,524 potential lincRNA clusters, of which the majority were previously unknown.

Stage and lineage specificity of lincRNAs
To determine the cellular specificity of lincRNA expression, we used pairwise comparison of protein-encoding mRNAs and lincRNAs in any two developmental stages of T cells to assess the difference in expression at one stage relative to expression at another stage. Overall mRNA expression was highly similar, but lincRNA expression exhibited substantial differences in any two T cell subsets generated in vivo (Fig. 1c) or obtained after in vitro differentiation (Fig. 1d). We further separated the T cell subsets into three groups (DN cells; DP cells, SP cells and T\(_{\text{reg}}\) cells; and naive CD4\(^+\) T cells and helper T cells) and determined the overlap of lincRNA expression or mRNA expression in each group (by Venn diagrams). The data indicated that 48–57% of lincRNAs were stage specific or lineage specific, in contrast to 6–8% of mRNAs (Fig. 1e–g). On the other hand, 75–80% of protein-encoding genes were shared by subsets in a group, in contrast to 13–16% of the lincRNA-encoding genes (Fig. 1e–g). To examine the expression of each individual lincRNA, we plotted a heat map of lincRNA expression in helper T cells. About 56% of lincRNAs (367) exhibited preferential expression in one T cell subset rather than in the others (Fig. 1h). Comparison of the T cell lincRNA ‘catalog’ generated here and the lincRNAs identified in mouse embryonic stem cells\(^{24}\) revealed that only 8.5% of the T cell lincRNAs (or 129 clusters) were also expressed in mouse embryonic stem cells (data not shown). In summary, we found that lincRNA expression was highly cell specific during the development and differentiation of T cells.

Most lincRNAs are polyadenylated and dynamically regulated
By comparing the data obtained by RNA-Seq analysis of total RNA and poly(A) RNA, we observed that some lincRNAs were polyadenylated in T cells (Fig. 2a). To determine which lincRNAs were polyadenylated and if the polyadenylated lincRNAs were dynamically regulated during the differentiation of helper T cells, we profiled poly(A) RNA by RNA-Seq at various time points (4 h, 8 h, 12 h, 24 h, 48 h, 72 h, 1 week and 2 weeks) during the differentiation of naive CD4\(^+\) T cells. Comparison of the expression of poly(A) RNAs, as calculated from the sequencing data, and total RNAs indicated a high correlation of LincR-Gata3-3′ expression in the two sets (Pearson correlation coefficient (\(r\) = 0.99)) across various subsets (Fig. 2b and Supplementary Table 2). Analysis of all lincRNA clusters revealed that over 50% of the lincRNAs expressed in helper T cells had an \(r\) value of >0.6 (Fig. 2c), which suggested that a large fraction of the lincRNAs were polyadenylated.

Analysis of polyadenylated lincRNAs at the time points during helper T cell differentiation assessed above indicated that the expression of many lincRNAs in naive CD4\(^+\) T cells was rapidly downregulated after 4 h of differentiation and then they were reexpressed after 48–72 h (Fig. 2d,e), which suggested that they might be involved in regulating T cell activation. We next used a heat map to visualize lincRNA expression at all those time points during the differentiation of naive CD4\(^+\) T cells into helper T cell subsets. Of the 539 lincRNAs, 19 were rapidly downregulated within 4 h of T cell differentiation and remained largely silenced throughout the later time points (for example, LincR-Chd2-5′-74K). One lincRNA preferentially expressed in T\(_{\text{H}2}\) cells, LincR-Slo-5′AS, was rapidly induced at 4 h, and its expression then gradually decreased during later differentiation (Fig. 2f). Most lincRNAs preferentially expressed in T\(_{\text{H}1}\) or T\(_{\text{H}2}\) cells exhibited substantial induction at 48–72 h and reached a plateau of expression within 1–2 weeks (for example, LincR-Gata3-3′ and LincR-Crr2-5′AS), while most lincRNAs preferentially expressed in T\(_{\text{H}1}\) cells were maximally induced within 48–72 h (Fig. 2f). Thus, most lincRNAs were polyadenylated and were dynamically regulated during T cell differentiation.

STAT4 activates lincRNAs preferentially expressed in T\(_{\text{H}1}\) cells
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 differentiation into T\(_{\text{H}1}\) and T\(_{\text{H}2}\) cells, respectively, by binding to and activating lineage-specific enhancers in these cells\(^{25}\). Using data obtained by ChIP-Seq analysis of STAT4 binding\(^{26}\), we found that STAT4 bound to 56% of the lincRNA-encoding genes (862; Supplementary Table 3), such as LincR-Gng2-5′-74K). STAT4 binding was stronger at lincRNA-encoding clusters preferentially expressed in T\(_{\text{H}1}\) cells than at other clusters (Fig. 3a). Deletion of STAT4 decreased expression of LincR-Gng2-5′ in T\(_{\text{H}1}\) cells, as assessed by comparison of wild-type and STAT4-deficient cells (Fig. 3a). In total, expression of 39% (90) of the lincRNA clusters preferentially expressed in T\(_{\text{H}1}\) cells was downregulated in STAT4-deficient T\(_{\text{H}1}\) cells, in contrast to 8% (83) of lincRNAs not preferentially expressed in T\(_{\text{H}1}\) cells (Fig. 3c). Because STAT4 may also repress lincRNAs not preferentially expressed in T\(_{\text{H}1}\) cells, we observed that the expression of more lincRNAs not preferentially expressed in T\(_{\text{H}1}\) cells than lincRNAs preferentially expressed in T\(_{\text{H}1}\) cells was upregulated in STAT4-deficient T\(_{\text{H}1}\) cells (Fig. 3c). Furthermore, lincRNAs with strong binding of STAT4 at the promoters of the genes encoding them, as measured in wild-type T\(_{\text{H}1}\) cells, were more likely to have downregulated expression in STAT4-deficient T\(_{\text{H}1}\) cells than were lincRNAs with weak STAT4 binding (Fig. 3d). Consistent with that, lincRNAs with stronger STAT4 binding were less likely to have upregulated expression in STAT4-deficient T\(_{\text{H}1}\) cells (Fig. 3d). LincRNA-encoding genes were generally less frequently bound by STAT4 than were protein-encoding genes (Supplementary Table 3). However, of the STAT4 targets, the frequency of positive or negative regulation was in general similar for lincRNA- and protein-encoding genes (Supplementary Table 3). In summary, our data indicated that STAT4 bound to and activated T\(_{\text{H}1}\)-specific lincRNA-encoding genes in T\(_{\text{H}1}\) cells.
STAT6 activates lincRNAs preferentially expressed in TH2 cells

To understand the function of STAT6 in regulating lincRNA expression in TH2 cells, we analyzed ChIP-Seq data to identify STAT6 binding to lincRNA-encoding genes in wild-type TH2 cells. We detected STAT6 binding in 56% of all lincRNA-encoding clusters (Supplementary Table 3), including LincR-Epas1-3'AS (Fig. 3e). The binding of STAT6 to promoters in TH2 cells was higher for lincRNA clusters preferentially expressed in TH2 cells than for other clusters (Fig. 3f). Expression of LincR-Epas1-3'AS was lower in TH2 cells from STAT6-deficient mice than in those from wild-type mice (Fig. 3e). In total, expression of 32% of the lincRNAs (56) preferentially expressed in TH2 cells was lower and expression of 12% of the other lincRNAs (131) was higher in STAT6-deficient TH2 cells than in wild-type TH2 cells (Fig. 3f). If both a lincRNA-encoding gene and a protein-encoding gene were activated or repressed by STAT6, they tended to be coregulated during T cell differentiation; we made a similar observation for STAT4-mediated gene regulation in TH1 cells (Supplementary Fig. 2). LincRNAs with strong binding of STAT6 at the promoters of the genes encoding them, as measured in wild-type TH2 cells, were more likely to have downregulated expression in STAT6-deficient TH2 cells than those with weak STAT6 binding (Fig. 3h). However, upregulation of lincRNA expression in STAT6-deficient TH2 cells was not significantly correlated with STAT6 binding in wild-type TH2 cells (Fig. 3h). In summary, STAT6 bound to and mediated the activation of TH2-specific lincRNA-encoding genes in TH2 cells.

Figure 2 Dynamic regulation of lincRNA expression during T cell differentiation

(a) Distribution of RNA-Seq ‘reads’ for poly(A) RNA and total RNA across the Linc-R-Gata3-3' cluster, in TH2 cells obtained after 2 weeks of polarization in vitro. (b) Expression of Linc-R-Gata3-3' by various helper T subsets (Supplementary Table 2), assessed by RNA-Seq analysis of poly(A) RNA and total RNA. (c) Inverse cumulative distribution of Pearson correlation coefficients for lincRNA expression assessed as in b. Random, simulation with no correlation assumed. (d) Quantification of lincRNAs expressed at various time points (left margin) during the differentiation from a naive CD4+ T cell (N) into various helper T cell subsets (top), assessed by RNA-Seq analysis of poly(A) RNA. w, weeks. (e) Heat map of gene expression at various time points (as in d) for lincRNAs expressed in naive CD4+ T cells; black rectangle outlines those first downregulated and then reactivated (assessed as in d). (f) Hierarchical cluster analysis of gene expression at various time points for all lincRNAs excluding those not expressed in any subset (as assessed in d); ‘X-pref’ indicates a lincRNA with maximal expression across all time points in lineage X 1.5-fold higher than the maximum from any other combination of lineage and time points (‘shared’ indicates all other conditions). Data are representative of two experiments (biological replicates).
RNA-Seq expression of LincR-Gng2-3′AS was lower in Tbx21−/− TH1 cells than in wild-type TH1 cells (Fig. 4d), consistent with a published report 14. In contrast, expression of LincR-Ccr2-5′AS (which was preferentially expressed in T12 cells) was modestly upregulated in Tbx21−/− TH1 cells relative to its expression in wild-type TH1 cells (Fig. 4d). Global analysis revealed that deletion of Tbx21 resulted in decreased expression of 54 lincRNAs and increased expression of 37 lincRNAs in TH1 cells (Fig. 4e). 36% (33 of 91) of the affected genes were preferentially expressed in T12 cells (data not shown). On the other hand, 16.3% (216 of those unaffected lincRNAs were preferentially expressed in T12 cells (P < 0.0001) (data not shown). Thus, our data indicated that T-bet bound to and contributed to the activation and repression of lincRNA-encoding genes in TH1 cells.

**GATA-3 regulates lincRNA expression in T12 cells**

GATA-3, a zinc-finger transcription factor, has high expression in TH2 cells and is critical in TH2 differentiation by regulating the expression of many T12 cell–specific genes 31. Analysis of published ChIP-Seq data of GATA-3 binding 32 showed that 28.5% (53) of the lincRNA-encoding gene clusters preferentially expressed in T12 cells were bound by GATA-3 (data not shown). In wild-type T12 cells, GATA-3 preferentially bound to the promoters of lincRNA-encoding genes that were preferentially expressed in T12 cells (Fig. 5a). Consistently, GATA-3-bound lincRNA-encoding gene clusters had higher expression than did clusters not bound by GATA-3 (Fig. 5b). To determine if GATA-3 contributed to the regulation of lincRNAs, we compared their expression in wild-type T12 and Gata3−/− T12 cells. GATA-3 deficiency resulted in much lower expression of LincR-Ccr2-5′AS (Fig. 5c). Global analysis of lincRNA expression revealed that GATA-3 deficiency resulted in lower expression of 30% (102) of lincRNA-encoding gene clusters bound by GATA-3, with 15% (149) of lincRNA-encoding gene clusters not bound by GATA-3 (Fig. 5d). Notably, 11% of lincRNA-encoding gene clusters bound by GATA-3 had higher expression in Gata3−/− T12 cells than in wild-type T12 cells, compared with 6% of the clusters not bound by GATA-3 (Fig. 5d).

To determine if GATA-3 coregulated the expression of proximal lincRNAs and protein-encoding genes, we separated lincRNAs into three groups (downregulated, upregulated and unchanged by GATA-3 deficiency) and examined the expression of protein-encoding genes within 100 kilobases (kb) of the lincRNA-encoding genes in T12 cells. About 37% (65 of 175) of downregulated lincRNA-encoding gene clusters had at least one gene nearby that also had lower expression (Fig. 5e); for lincRNA-encoding genes with no change or upregulation in expression in Gata3−/− versus wild-type T12 cells, the frequency of nearby genes with the same ‘direction’ in expression change was 25.3% (187 of 740) or 8.5% (6 of 71), respectively (Fig. 5e). On the other hand, lincRNA-encoding genes upregulated in Gata3−/− T12 cells were more likely to have a nearby protein-encoding gene that was also upregulated (Fig. 5e). These results indicated that a substantial fraction of lincRNA-encoding genes were coregulated with their neighboring protein-encoding genes, which suggested that the protein-encoding and lincRNA-encoding genes might share a GATA-3-responsive enhancer.

**Inference with lincRNA function during T cell differentiation**

Because several highly inducible lincRNA-encoding genes in helper T cells were adjacent to genes encoding proteins critical to T cell function, such as the LincR-Gata3-3′ cluster (located 3′ of Gata3) and the LincR-Ccr2-5′AS cluster (located between the chemokine receptor–encoding genes Ccr3 and Ccr2), we examined the position of all the lincRNA-encoding genes identified relative to that of neighboring protein-encoding genes (within 100 kb). Many lincRNA-encoding genes were adjacent to genes encoding proteins in groups highly enriched for immunological functions, as...
Figure 4  T-bet regulates lincRNA expression in T\(_H\)1 cells. (a,b) Distribution of RNA-Seq ‘reads’ for total RNA across the LincR-\(\Delta\)g3-AS cluster (a) or LincR-Ccr2-5’AS cluster (b) in naive CD4\(^+\) T cells and other helper T cells, and ChiP-Seq peaks (arrowheads) for T-bet in the same region in wild-type cells. (c) Cumulative distribution of ChiP-Seq ‘read’ density for T-bet at promoters of genes encoding lincRNAs preferentially expressed in wild-type T\(_H\)1 cells (249; defined as in Fig. 1h) and not. *P < 0.01 (Kolmogorov-Smirnov test). (d) Distribution of RNA-Seq ‘reads’ for total RNA across the LincR-\(\Delta\)g3-AS cluster (upper) or LincR-Ccr2-5’AS cluster (lower) in wild-type and Tbx21\(^{-/-}\) T\(_H\)1 cells after 2 weeks of polarization. FDR, false-discovery rate. (e) Ratio of expression in Tbx21\(^{-/-}\) T\(_H\)1 cells that in wild-type cells (vertical axis) and average expression of lincRNA in Tbx21\(^{-/-}\) T\(_H\)1 cells versus that in wild-type T\(_H\)1 cells after 2 weeks of polarization (horizontal axis), presented as a Bland-Altman plot; background, blurring of values for protein-encoding genes; red, lincRNA with a significant change in expression of over 1.5-fold (false-discovery rate, <0.05). Data are representative of two experiments (biological replicates).

defined by enrichment analysis with the KEGG database (Kyoto Encyclopedia of Genes and Genomes) and the Gene Ontology (GO) project (Supplementary Table 4). This suggested possible coevolution of lincRNA- and protein-encoding genes for the control of specialized functions.

To identify potential functions for lincRNAs in T cells, we analyzed the coexpression of lincRNA- and protein-encoding genes during differentiation from naive CD4\(^+\) T cells into various helper T cell subsets (T\(_H\)1, T\(_H\)2, T\(_H\)17 and iTreg), each including eight time points (defined in Supplementary Fig. 1a; Supplementary Fig. 3a). Genome wide, protein-encoding genes that were positively correlated with a lincRNA in terms of expression tended to be located near that lincRNA-encoding gene (Supplementary Fig. 3b). Many lincRNA-encoding genes were coexpressed with genes encoding proteins in groups highly enriched for terms of the GO project related to immunological and/or defense response, regulation of T cell-mediated cytotoxicity, and ribosome biogenesis and cell-cycle regulation (Supplementary Fig. 3c). We next examined the expression of the 151 lincRNAs associated with the genes encoding proteins involved in ribosome biogenesis during the differentiation from naive CD4\(^+\) T cells into distinct helper T cell lineages. We found that 31 of the 151 lincRNAs underwent transient induction after 4 h, followed by repression after 24 h, while most of the remaining lincRNAs were transiently repressed and then were substantially upregulated by 72 h (Supplementary Fig. 4a). In contrast, 87% of genes encoding proteins involved in ribosomal biogenesis were rapidly induced at 4 h and then repressed after 24 h (Supplementary Fig. 4b). In summary, our data indicated that many lincRNAs were coexpressed with genes encoding proteins involved in immunological function, which suggested a regulatory role for lincRNAs in T cell differentiation and function; the dynamic regulation of the lincRNA-encoding genes correlated with that of genes encoding proteins involved in ribosome biogenesis suggested that many of these lincRNAs may restrict ribosomal functions for cells in the resting state.

Figure 5  GATA-3 regulates expression of lincRNAs in T\(_H\)1 cells. (a) Cumulative distribution of ChiP-Seq ‘read’ density for GATA-3 at promoters of genes encoding lincRNAs preferentially expressed in wild-type T\(_H\)1 cells (186; defined as in Fig. 1h) or not. *P < 0.01 (Kolmogorov-Smirnov test). (b) Cumulative distribution of the expression of lincRNAs bound by GATA-3 (+GATA-3) or unbound (–GATA-3) at promoters in wild-type T\(_H\)2 cells after 2 weeks of polarization. *P < 0.01 (Kolmogorov-Smirnov test). (c) Distribution of RNA-Seq ‘reads’ for total RNA across the LincR-Ccr2-5’AS cluster in Gata3\(^{-/-}\) and wild-type T\(_H\)2 cells (top and middle) and ChiP-Seq peaks (arrowheads) for GATA-3 in wild-type cells in the same region (bottom). (d) Frequency of lincRNAs downregulated (left) or upregulated (right) in Gata3\(^{-/-}\) T\(_H\)2 cells among lincRNAs bound by GATA-3 or unbound at promoters. *P < 0.0001, **P = 0.005 (\(\chi^2\)-test). (e) Frequency of lincRNA-encoding genes with at least one nearby protein-encoding gene (within 100,000 bp) downregulated (left) or upregulated (right) after deletion of Gata3, among three groups of lincRNAs sorted on the basis of their response to GATA-3 deficiency in T\(_H\)2 cells (downregulated (175), upregulated (71) or unchanged); lincRNAs around which the surrounding 100,000–base pair region contained both upregulated and downregulated genes or no genes were excluded. *P < 0.01 (\(\chi^2\)-test). Data are from one experiment with two independent pools of cells.
Figure 6 LincR-Ccr2-5′-AS regulates gene expression and the migration of TH2 cells. (a) Distribution of total RNA ‘reads’ for Ccr1, Ccr3, LincR-Ccr2-5′-AS, Ccr2 and Ccr5 in TH2 cells treated with shLuc (control) or with sh36 or sh40 (LincR-Ccr2-5′-AS knockdown), and the distribution of H3K4me3 ‘reads’ in the same region in cells not treated with shRNA. Inset, enlargement of the top three rows across LincR-Ccr2-5′-AS (different scales for vertical axes). (b) Migration efficiency of CD45.2+ TH2 cells transfected with shLuc (n = 5), sh36 (n = 4) or sh40 (n = 4) recovered from the lungs of the recipient C57BL/6 mice 20 h after cotransfer with CD45.1+ TH2 cells not treated with shRNA, presented as the ratio of shRNA-treated cells to cells not treated with shRNA, with the mean for the shLuc-treated group set as 1. Each symbol represents an individual mouse; small horizontal lines indicate the mean (± s.e.m.). *P < 0.01 (one-tailed t-test). (c) All protein-encoding genes (left) or those preferentially expressed in wild-type TH2 cells (right; defined as in Fig. 1h) upregulated, downregulated and unchanged by depletion of LincR-Ccr2-5′-AS. (d) Pearson correlation coefficient (r) for the expression of LincR-Ccr2-5′-AS versus protein-encoding genes sorted by their response to depletion of LincR-Ccr2-5′-AS in TH2 cells: upregulated (709), downregulated (656) or unchanged (9,329). *P < 0.01 (one tailed t-test). (e) Frequency of genes protein-encoding genes (mRNA) downregulated (left) or upregulated (right) in response to depletion of LincR-Ccr2-5′-AS, sorted by similarity in gene expression (r) to LincR-Ccr2-5′-AS: r > 0.6 (119), r < –0.6 (100) and others. ND, not determined; NS, not significant; *P < 0.01 (χ2-test). Data are from one experiment.

LincR-Ccr2-5′-AS regulates the migration of TH2 cells to the lungs

The groups of genes with expression patterns associated with LincR-Ccr2-5′-AS showed enrichment for genes encoding proteins involved in chemokine-mediated signaling pathways (Supplementary Fig. 3c): 7 of the 23 genes annotated in this pathway had expression that highly correlated with LincR-Ccr2-5′-AS expression, while 6 of the 23 were located in the same genomic region as the gene encoding LincR-Ccr2-5′-AS (Supplementary Table 5), which suggested coregulation of their expression. To directly determine if LincR-Ccr2-5′-AS controlled the chemokine-mediated migration of T cells, we designed shRNA to knock down its expression in TH2 cells. LincR-Ccr2-5′-AS was transcribed in a direction opposite to the transcription of Ccr2 (Supplementary Fig. 5a) and was expressed specifically in TH2 cells (Supplementary Fig. 5b). Two independent shRNAs specifically targeting LincR-Ccr2-5′-AS (sh36 and sh40) efficiently knocked down LincR-Ccr2-5′-AS expression in TH2 cells (Fig. 6a, inset). TH2 cells infected with lentivirus containing sh36 or sh40 had production of interleukin 4 (IL-4) similar to that of cells infected with lentivirus containing the control shRNA shLuc (data not shown), which suggested that LincR-Ccr2-5′-AS did not regulate IL-4 production in TH2 cells. However, knockdown of LincR-Ccr2-5′-AS expression by sh36 or sh40 resulted in lower expression (from 1.5- to 12-fold) of Ccr2, Ccr3 and Ccr5 than in cells treated with shLuc; all these genes are located in the vicinity of the gene encoding LincR-Ccr2-5′-AS (Fig. 6a and Supplementary Table 6).

Because the trafficking of TH2 cells into the lungs is dependent on signaling via receptors for chemotactic factors22, we assessed the ability of TH2 cells in which LincR-Ccr2-5′-AS was knocked down to migrate to the lungs in vivo. We transduced CD45.2+ TH2 cells with shLuc, sh36 or sh40, mixed the successfully transduced CD45.2+ TH2 cells with congenic CD45.1+ TH2 cells and transferred the mixture into naive C57BL/6 mice. The efficiency with which LincR-Ccr2-5′-AS-deficient TH2 cells migrated to the lungs 20 h after transfer was significantly impaired (Fig. 6b). These results indicated that LincR-Ccr2-5′-AS contributed to the migration of TH2 cells, which correlated with its ability to modulate the expression of several chemokine receptors.

RNA-Seq analysis of TH2 cells transfected with LincR-Ccr2-5′-AS-specific shRNA showed that the expression of 709 and 656 mRNAs was significantly upregulated and downregulated, respectively, relative to their expression in TH2 cells transfected with shLuc (change in expression, >1.5-fold; false-discovery rate, <0.05) and that mRNAs preferentially expressed in TH2 cells were three times more likely to be downregulated than were other mRNAs (Fig. 6c and Supplementary Table 7). The genes downregulated by depletion of LincR-Ccr2-5′-AS encoded proteins that as a group showed enrichment for biological processes such as cell cycle and nuclear division, whereas the genes upregulated encoded proteins with enrichment for regulation of immune-system processes and defense responses (Supplementary Table 8). Transduction of sh36 or sh40 into cells cultured under TH17 conditions did not yield significant changes in gene expression (data not shown), which correlated with the lack of expression of LincR-Ccr2-5′-AS in TH17 cells and indicated that off-target effects were minimal.

We further investigated the relationship between GATA-3 and LincR-Ccr2-5′-AS and their downstream target protein-encoding genes. We found that 170 genes were activated by both GATA-3 and LincR-Ccr2-5′-AS and their downstream target protein-encoding genes. We found that 170 genes were activated by both GATA-3 and LincR-Ccr2-5′-AS, 122 genes were repressed by both, 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 (Supplementary Fig. 6a). Regardless of the ‘direction’ of regulation, the genes affected by both LincR-Ccr2-5′-AS and GATA-3 were significantly over-represented relative to those affected by only one, and as a group, their shared target genes showed considerable enrichment for TH2-specific genes (Supplementary Fig. 6b). GO-term-enrichment analysis for genes responsive to both knockdown of LincR-Ccr2-5′-AS and deletion of Gata3 revealed that GO terms related to cell cycle and immune function were among the top categories (Supplementary Fig. 6c and Supplementary Table 9), which further indicated that LincR-Ccr2-5′-AS together with GATA-3 was a critical regulator of T cell differentiation and immunological function.

LincR-Ccr2-5′-AS regulates coexpressed protein-encoding genes

When we analyzed the Pearson correlation coefficient of expression for LincR-Ccr2-5′-AS and the genes upregulated, downregulated
or unchanged after transduction of LincR-Ccr2-5′AS-specific shRNA in T112 cells, we found that expression of the downregulated group showed significantly higher correlation with LincR-Ccr2-5′AS expression during the differentiation of naive CD4+ T cells into various helper T cells than did the unchanged and upregulated groups (Fig. 6d). Analysis of the frequency of protein-encoding genes downregulated after the transduction of LincR-Ccr2-5′AS-specific shRNA, among various groups of genes coexpressed with LincR-Ccr2-5′AS in T112 cells, revealed that genes positively correlated with LincR-Ccr2-5′AS were more likely to be downregulated than were other genes in the absence of LincR-Ccr2-5′AS, while comparison of the frequency of protein-encoding genes upregulated after the shRNA-mediated knockdown indicated that genes negatively correlated with LincR-Ccr2-5′AS were not preferentially upregulated in the absence of LincR-Ccr2-5′AS (Fig. 6e). These results suggested that LincR-Ccr2-5′AS positively regulated many coexpressed protein-encoding genes.

LincRNAs may regulate gene expression through modulation of chromatin structure at target sites or may function as enhancer RNAs3. To determine whether LincR-Ccr2-5′AS regulates chromatin state, we used sequencing of DNase I–hypersensitive sites and ChIP followed by quantitative PCR to assess chromatin accessibility and binding of RNA polymerase II near the Ccr2 and Ccr3 loci in T112 cells transduced with sh36, sh40 or shLuc. Although the expression of Ccr2 and Ccr3 was substantially decreased, we detected no substantial change in chromatin accessibility or the modification of histone H3 trimethylated at Lys4 (H3K4me3) or binding of RNA polymerase II in cells transduced with LincR-Ccr2-5′AS-specific shRNA compared with that in cells transduced with shLuc (data not shown), which suggested that LincR-Ccr2-5′AS regulated the expression of Ccr2 and Ccr3 via a mechanism distinct from modulation of chromatin accessibility or recruitment of RNA polymerase II.

DISCUSSION

In this study, we identified 1,524 genomic regions expressing lincRNAs in 42 samples from T cells at various developmental and differentiation stages and found that the lincRNAs were highly stage specific or lineage specific, consistent with the proposal that they are important regulators of the development, differentiation and function of T cells. Among the 1,524 lincRNA clusters, only one lincRNA, TMEVPG1 (LincR-Ifng-3′AS), has been studied in T cells and is reported to have an important role in controlling infection with Theiler’s virus13, while no functions have been reported for any other lincRNAs in T cells, to our knowledge. It has been reported that the functions of certain lincRNAs in a particular pathway could be inferred from data indicating their coexpression with protein-encoding genes3, including the function of lincRNA-p21 in apoptosis mediated by the tumor suppressor p53 (ref. 33) and that of lincRNA-ROR in maintaining the pluripotency of embryonic stem cells34. Thus, our observation that the expression of many lincRNA-encoding genes showed high correlation with that of genes encoding protein associated with RNA processing and ribosome biogenesis, cell cycle and immune responses suggested that those lincRNAs may be functionally involved in those biological processes in T cells. In particular, expression of the gene encoding the T112-specific LincR-Ccr2-5′AS showed high correlation with that of genes encoding proteins involved in the chemokine signaling pathway in T112 cells. Knockdown of LincR-Ccr2-5′AS decreased the expression of its neighboring chemokine receptor–encoding genes and compromised the migration of T112 cells to the lung tissues, which revealed a critical function for this lincRNA in T112 cell function and confirmed the validity of inferring the function of lincRNAs from their coexpression with protein-encoding genes. Thus, future studies should assess the function of other lincRNAs by loss-of-function or gain-of-function assays.

LincRNAs may regulate gene expression via different mechanisms, including acting as enhancer RNA9, repressing microRNA targeting10,11 and binding to target genes to recruit chromatin-modifying enzymes7,8. TMEVPG1 (LincR-Ifng-3′AS)acts together with T-bet to mediate transcription of Ifng14, probably through direct interaction with the MLL complexes to facilitate the methylation of H3K4 at its target sites15. LincR-Gata3-3′ is located near the T cell–specific enhancer of GATA-3 expression15 and therefore it could act as an enhancer RNA. Further experiments are needed to determine the function of LincR-Gata3-3′. LincR-Ccr2-5′AS was required for efficient expression of the nearby genes Ccr2 and Ccr3, which suggested that it could activate those genes in cis. However, H3K4me3 modification, DNase accessibility and binding of RNA polymerase II to Ccr2 and Ccr3 were not affected by knockdown of LincR-Ccr2-5′AS, which suggested that it does not function to recruit histone-modifying enzymes or to modify the chromatin structure of those loci. Thus, LincR-Ccr2-5′AS may regulate the expression of its target genes after the initiation of transcription. Because LincR-Ccr2-5′AS also affects global gene expression, it may additionally act in trans.

Published analysis of lincRNAs in various human organs has indicated that lincRNA expression is more tissue specific than is mRNA expression1. In agreement with that, we found that lincRNA expression was highly stage specific and cell specific during T cell differentiation. In helper T cells, cell-specific lincRNAs constituted 10–40% of the total lincRNAs detected in a cell type, which suggested that lincRNA expression was tightly regulated during differentiation. Our data indicated that key transcription factors, including T-bet and STAT4 for the T111 lineage, and GATA-3 and STAT6 for the T112 lineage, were largely accountable for the lineage-specific expression of T cell lincRNAs. Since both lincRNA- and protein-encoding genes were subjected to a similar degree of positive or negative regulation by those transcription factors, similar mechanisms may be used by those factors in regulating both coding and noncoding genes. On the other hand, lincRNAs may affect the expression of neighboring protein-encoding genes to reinforce or attenuate gene regulation by transcription factors, which would add another layer to the regulatory network of transcription program that underlies T cell development and differentiation. Our data set should serve as a resource for the study of transcriptional regulatory networks during T cell development and differentiation by comparison of the dynamic expression of genes encoding proteins, including transcription factors, cell surface markers and signaling molecules, with that of lincRNA-encoding genes. We expect that further characterization of the lincRNAs identified in this study will reveal important functions of lincRNAs in the development, differentiation and immune responses of T cells.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. GEO: raw sequence data and processed files, GSE48138.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the DNA Sequencing Core facility of the National Heart, Lung and Blood Institute (NHLBI) for sequencing the ChIP-Seq and RNA-Seq libraries; J. Edwards for
most cell-sorting experiments; the flow cytometry core of NHLBI for some cell-sorting experiments and analysis; H. Cao for comments on the knockdown of lincRNA by shRNA; D. Northrup for critical reading and editing of the manuscript; H. Zhang for sharing experience with chemokines and chemokine receptors; P. Burr for RNA-Seq; and X. Zheng for sharing code for binomial tests. This study used the Biowulf Linux cluster of the US National Institutes of Health. Supported by the Division of Intramural Research of the NHLBI and NIAID (US National Institutes of Health).

AUTHOR CONTRIBUTIONS
G.H., J.Z. and K.Z. conceived of the study, designed experiments and data analysis, and wrote the manuscript; Q.T., S.S. and F.Y. did experiments and edited the manuscript; G.H. analyzed the data; and T.M.E. and S.A.M. contributed RNA-Seq data for STAT4-deficient T$_{H}$1 cells, STAT6-deficient T$_{H}$2 cells and the corresponding wild-type T$_{H}$1 and T$_{H}$2 cells.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Cabili, M.N. et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25, 1915–1927 (2011).
2. Derrien, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 22, 1775–1789 (2012).
3. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. 81, 145–166 (2012).
4. Spizzo, R., Almeida, M.I., Colombatti, A. & Calin, G.A. Long non-coding RNAs and cancer: a new frontier of translational research. Oncogene 31, 4577–4587 (2012).
5. Payer, B. & Lee, J.T. X chromosome dosage compensation: how mammals keep the balance. Annu. Rev. Genet. 42, 733–772 (2008).
6. Rinn, J.L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129, 1311–1323 (2007).
7. Khalil, A.M. et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. USA 106, 11667–11672 (2009).
8. Zhao, J. et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol. Cell 40, 939–953 (2010).
9. Orom, U.A. et al. Long noncoding RNAs with enhancer-like function in human cells. Cell 143, 46–58 (2010).
10. Cesana, M. et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell 147, 358–369 (2011).
11. Hansen, T.B. et al. Natural RNA circles function as efficient microRNA sponges. Nature 495, 384–388 (2013).
12. Zhu, J., Yamane, H. & Paul, W.E. Differentiation of effector CD4 T cell populations. Annu. Rev. Immunol. 28, 445–489 (2010).
13. Vigneau, S., Rohrich, P.S., Brahic, M. & Bureau, J.F. Tmevpg1, a candidate gene for the control of Th1/Th2 dominance in mice, could be implicated in the regulation of IFN-γ expression in the livers of Th1 and Th2 cells. J. Virol. 77, 5632–5638 (2003).
14. Collier, S.P., Collins, P.L., Williams, C.L., Boothby, M.R. & Aune, T.M. Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of IFNγ by Th1 cells. J. Immunol. 189, 2084–2088 (2012).
15. Gomez, J.A. et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-γ locus. Cell 152, 743–754 (2013).
16. Pang, K.C. et al. Genome-wide identification of long noncoding RNAs in CD8+ T cells. J. Immunol. 182, 7738–7748 (2009).
17. Pagani, M. et al. Role of microRNAs and long-non-coding RNAs in CD4+ T-cell differentiation. Immunol. Rev. 253, 82–96 (2013).
18. Guttman, M. et al. Chromatin signature reveals over a thousand highly conserved long non-coding RNAs in mammals. Nature 458, 223–227 (2009).
19. Pruitt, K.D., Tatusova, T., Brown, G.R. & Maglott, D.R. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. 40, D130–D135 (2012).
20. Flicek, P. et al. Ensembl 2013. Nucleic Acids Res. 41, D84–D85 (2013).
21. Meyer, L.R. et al. The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res. 41, D64–D69 (2013).
22. Bu, D. et al. NONCODE v3.0: integrative annotation of long noncoding RNAs. Nucleic Acids Res. 40, D210–D215 (2012).
23. Kong, L. et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35, W345–W349 (2007).
24. Sigova, A.A. et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc. Natl. Acad. Sci. USA 110, 2876–2881 (2013).
25. Vahedi, G. et al. STATA5s shape the active enhancer landscape of T cell populations. Cell 151, 981–993 (2012).
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 36, 840–851 (2010).
27. Szabo, S.J. et al. A novel transcription factor, T-bet, directs Th1 lineage commitment. Cell 100, 655–669 (2000).
28. Zheng, W. & Flavell, R.A. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. Cell 89, 587–596 (1997).
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 37, 660–673 (2012).
30. Yagi, R., Zhu, J. & Paul, W.E. An updated view on transcription factor GATA-3-mediated regulation of Th1 and Th2 cell differentiation. Int. Immunol. 23, 415–420 (2011).
31. Wei, G. et al. Genome-wide analyses of transcription factor GATA-3-mediated gene regulation in distinct T cell types. Immunity 35, 299–311 (2011).
32. Mathew, A., Medoff, B.D., Carafone, A.D. & Luster, A.D. Cutting edge: Th2 cell trafficking into the allergic lung is dependent on chemotactic protein–receptor signaling. J. Immunol. 169, 651–655 (2002).
33. Huarte, M. et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell 142, 409–419 (2010).
34. Loewer, S. et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat. Genet. 42, 1113–1117 (2010).
35. Hosoya-Omura, S. et al. An NK and T cell enhancer lies 280 kilobase pairs 3′ to the gata3 structural gene. Mol. Cell. Biol. 31, 1894–1904 (2011).
ONLINE METHODS

Mice. C57BL/6 mice were from the Jackson Laboratory. T-bet-deficient mice carrying the T-bet-ZsGreen reporter and their wild-type control counterparts have been described\(^\text{35}\). Mice withloxP-flankedGata3alleles on the C57BL/6 background were described previously\(^\text{36}\). Mice expressing a green fluorescent protein (GFP) reporter forFoxp3(Foxp3\textsuperscript{GFP})mice were from Taconic Farms.

Wild-type, STAT4-deficient and STAT6-deficient mice of the BALB/c background were from the Jackson Laboratory. Mice were used at 8–12 weeks of age. All mice were maintained under specific pathogen–free conditions and were treated by an animal study protocol approved by the Animal Care and Use Committee of NIAID.

T cell isolation and differentiation. Single-cell suspensions from thymi of C57BL/6 mice were stained with fluorescein isothiocyanate–anti–CD4 (GK1.5; eBioscience), allophycocyanin-indocarbocyanine–anti–CD4 (IM7; eBioscience), allophycocyanin–anti–CD3 (145-2C11; eBioscience), Pacific blue–anti–CD8 (53-6.7; eBioscience), phycoerythrin–anti–CD25 (PC61.5; eBioscience) and peridinin chlorophyll protein–cyanine 5.5–anti–CD69 (1H.2F3; eBioscience) and were sorted as CD4\textsuperscript{+}CD8\textsuperscript{−}CD44\textsuperscript{+}CD25\textsuperscript{−} cells (DN1), CD4\textsuperscript{+}CD8\textsuperscript{+}CD44\textsuperscript{+}CD25\textsuperscript{−} cells (DN2), CD4\textsuperscript{+}CD8\textsuperscript{−}CD34\textsuperscript{−}CD44\textsuperscript{−}CD25\textsuperscript{−} cells (DN3), CD4\textsuperscript{+}CD8\textsuperscript{−}CD34\textsuperscript{−}CD44\textsuperscript{−}CD25\textsuperscript{−} cells (DN4), CD4\textsuperscript{+}CD8\textsuperscript{−}CD69\textsuperscript{+} cells (DP1), CD4\textsuperscript{+}CD8\textsuperscript{−}CD69\textsuperscript{−} cells (DP2), CD3\textsuperscript{−}CD4\textsuperscript{+}CD8\textsuperscript{−} cells (CD4\textsuperscript{+} SP) and CD3\textsuperscript{−}CD4\textsuperscript{−}CD8\textsuperscript{+} cells (CD8\textsuperscript{+} SP) with a FACSAria (BD Biosciences). T\textsuperscript{+}reg cells were isolated from the lymph nodes of Foxp3\textsuperscript{GFP}mice. Cells were stained with allophycocyanin–anti–CD4 (RM4-5; eBioscience) and were sorted as CD4\textsuperscript{+} GFP\textsuperscript{+} populations with a FACSaria (BD Biosciences).

CD4\textsuperscript{−} T cells were isolated from the lymph nodes by negative selection\(^\text{37}\). For purification of naive CD4\textsuperscript{+} T cells, lymph node cells were stained with Pacific Blue–anti–CD62L (MEL-14; eBioscience), allophycocyanin–anti–CD4 (identified above), allophycocyanin–indocarbocyanine–anti–CD4 (identified above) and phycoerythrin–anti–CD25 (identified above) and were sorted as CD4\textsuperscript{+}CD8\textsuperscript{−}CD44\textsuperscript{+}CD25\textsuperscript{−} cells with a FACSaria (BD Biosciences). For depletion of T cells among antigen–presenting cells, spleen cells were incubated for 45 min at 37 °C with anti–Thy-1.2 (1-20C11; prepared in-house) and rabbit complement (Cedarlane Laboratories Limited). Cells were then irradiated at 30 Gy (3000 rads). For most experiments, CD4\textsuperscript{+} T cells were cultured for 3 d together with antigen–presenting cells (at a ratio of 1:10) in complete RPMI-1640 medium (RPMI-1640 (21800-024; Life Technologies)) supplemented with 10% FCS, 2-mercaptoethanol, glutamine, penicillin, streptomycin and sodium pyruvate in the presence of anti–CD3 (1 µg/ml; 2C11; production service provided by Harlan) and anti–CD28 (11453D; Life Technologies) along with the appropriate combinations of antibodies and cytokines as follows: T\textsuperscript{H1} conditions, IL-12 (5 ng/ml), anti–IFN-γ (10 µg/ml; XMG1.2; production service provided by Harlan) and anti–IL-10 (10 µg/ml; 10D.11; Peptech), IL-2 (50 U/ml; provided by Biological Resource Branch, National Cancer Institute) and anti–IL-4 (10 µg/ml; 11B1; provided by Biological Resource Branch, National Cancer Institute); T\textsuperscript{H2} conditions, IL-4 (5 000 U/ml; 214-14; Peptech), IL-6 (10 ng/ml; 1011B-1; Peptech), IL-7 (50 U/ml; production service provided by Harlan) and psPAX2 packaging plasmids were cotransfected into 293T cells for packaging of lentivirus particles. Naive CD4\textsuperscript{+} T cells were infected with the lentiviral vectors containing shluc, sh36 or sh40 and then cultured under TH2 conditions. Cells were transferred to complete RPMI-1640 medium containing puromycin (5 µg/ml) and IL-7 (1 ng/ml) for 48 h of culture. Populations of cells were further expanded for 3–4 d under T\textsuperscript{H2} polarization conditions. Cells were transfected to complete RPMI-1640 medium containing puromycin (5 µg/ml) and IL-7 (1 ng/ml) for 48 h of culture. Populations of cells were further expanded for 3–4 d under T\textsuperscript{H2} polarization conditions.

RNA-Seq analysis of total RNA. Total RNA was purified with an mirNeasy micro kit (217084; Qiagen) and a DNase set (79254; Qiagen) for the option of on-column digestion with DNase. 10 ng purified total RNA was used for amplification of cDNA with the Ovation RNA-Seq System V2 (7102-08; NuGEN Technologies). 200 ng cDNA was sonicated in a Diagenode Bioruptor (level M, for a total of 30 min of 20 s on and 20 s off) to size range of 100–400 bp. Indexed libraries were prepared with a Multiplexing Sample Preparation Oligonucleotide Kit (1005709; Illumina) and an End-It DNA End-repair Kit (ER81050; Epitope) according to the user’s manual (Epitope) and sample preparation guide (Illumina).

Fractionation of nuclear and cytosolic RNA and strand–specific RNA-Seq. T\textsuperscript{H2} cells (5 × 10\textsuperscript{6}) differentiated for 2 weeks were harvested and washed with 1× PBS. The cell pellet was carefully resuspended in 500 µl ice-cold Qiagen RN L buffer (50 mM Tris-Cl, pH 8.0, 140 mM NaCl, 1.5 mM MgCl\textsubscript{2} and 0.5% (vol/vol) NP-40). After incubation on ice for 5 min, samples were centrifuged at 3,000 r.p.m. for 5 min at 4 °C. Supernatants containing the cytosolic fraction were mixed with seven volumes of Qiazol. Pellets were washed twice with PBS plus 1 mM EDTA, and washed pellets containing nuclei were lysed with 700 µl Qiazol. RNA was purified with an mirNeasy mini kit (Qiagen). Cytosolic sample had to be loaded onto one column repeatedly. Then, 900 ng purified RNA was subjected to library preparation with TrueSeq stranded Total RNA LT Sample Prep Kit–set A (RS-122-2201; Illumina) according to the user’s manual, with 13 cycles for the final PCR step.

Adaptive transfer and flow cytometry. Six-week-old female C57BL/6 (CD45.2\textsuperscript{+}) mice and B6.SJL-Cd45a(Ly5a)/Nai (CD45.1\textsuperscript{+}; line 7) mice were from Taconic and the Taconic NIAID repository, respectively. Naive CD4\textsuperscript{+} T cells from C57BL/6 mice were purified by cell sorting, then were infected with lentiviruses containing shluc, sh36 or sh40 and then cultured under T\textsuperscript{H2} conditions for 2 rounds, followed by cell sorting for GFP\textsuperscript{+} cells (where GFP expression indicates successful transduction). 2 × 10\textsuperscript{5} CD45.2\textsuperscript{+}T\textsuperscript{H2}cells were mixed with 0.5 × 10\textsuperscript{5} CD45.1\textsuperscript{+} T\textsuperscript{H2} cells and the mixture was injected intravenously into C57BL/6 mice. Twenty hours later, the migration of transferred T\textsuperscript{H2} cells into the lungs was assessed by flow cytometry. Mice were killed and their lungs were immediately perfused with 5 ml PBS. Lungs were removed and then minced with scissors to a fine slurry in 10 ml digestion buffer per lung (RPMI-1640 medium containing 5 µg/ml liberase TL (05401020001; Roche Diagnostics) and 5 U/ml DNase (04562820001; Roche Diagnostics)), followed by enzymatic digestion for 30 min at 37 °C. Digested lung tissue was transferred to a 40-µm cell strainer on a 50-ml tube and was pushed through the sieve with the syringe plunger. The total lung cell suspension was pelleted and then was suspended for 30 s in 1 ml ACK lysing buffer (A10492-01; Invitrogen). The reaction was stopped by the addition of 10 ml 1× Hanks’ balanced-salt solution.
with 3% FBS. The pellet was resuspended in the same medium and stained on ice for 30 min with a ‘cocktail’ of dye (Fixable Viability Dye eFluor780; ebioscience) and antibodies (eFluor 450–anti-CD4 (RM4-5; ebioscience), phycoerythrin–anti-CD45.1 (A20; ebioscience), allophycocyanin–anti-CD45.2 (104; ebioscience) and anti-FeCyI/III (2.4G2; production service provided by Harlan)). Data were collected with an LSR II (BD) and analyzed with FlowJo software (Tree star).

Data analysis. Sequence alignment and RNA-Seq ‘read’–enriched regions. Sequence ‘reads’ were mapped to the mouse genome (the mm9 National Center for Biotechnology Information assembly of the mouse genome) by Bowtie software for the alignment of short DNA sequences with default settings39. ‘Reads’ mapped to multiple positions were discarded. RNA-Seq ‘read’–enriched regions (or islands) were identified by SICER (spatial clustering approach for the identification of ChIP-enriched regions)40 (window size = 100 bp, gap size = 200 bp, 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 did not overlap any genomic region annotated by the reference sequence database (RefSeq) of the National Center for Biotechnology Information15, the Ensemble project of genome databases20 and the UCSC Genome Browser21, and if it did not overlap any transcript (assembled from our samples) extended from those annotated genomic regions. RNA-Seq islands in the same intergenic regions were clustered on the basis of similarity in expression profiles across all RNA-Seq libraries as follows: island i joined cluster C if the Pearson correlation coefficient of gene expressions between i and at least one member from cluster C was >0.8, and if i was the nearest island to cluster C; the 5’-most island was chosen as a ‘seed’ with which to initiate the clustering, and i constituted a ‘seed’ of a new cluster if it did not satisfy the two conditions. A compilation of lincRNA clusters in GTF format and the distributions of RNA-Seq ‘reads’ in BEDGraph format (GEO accession code, GSE48138) can be directly uploaded to the UCSC Genome Browser for visualization (Supplementary Fig. 7).

Promoter definition of lincRNAs. Genomic regions enriched for H3K4me3 signals were chosen as a proxy for potential promoters for lincRNAs. We collected publicly available ChIP-Seq data sets for H3K4me3 from the GEO (Gene Expression Omnibus) database repository for DN cells, DP cells, CD8+ T cells, Treg cells, naive T cells, Treg1 cells, Treg2 cells, Treg17 cells and iTreg cells26,31,41–43. ChIP-Seq ‘read’–enriched regions were identified by SICER40 (window size = 100 bp, gap size = 200 bp, E value = 0.1).

The nomenclature system. Each lincRNA cluster was named according to the nomenclature style ‘LinCR-RefGen-5′ (or 3′)S(or AS)-Dis’, where ‘RefGen’ means ‘reference gene’ (gene symbol) and ‘Dis’ indicates distance from the reference gene. For each cluster, we identified on each side the nearest protein-encoding gene and chose the one more similar to the cluster in terms of expression profile (measured by Pearson correlation coefficient) of the reference gene. For clusters located downstream of the reference gene, a tag 3′ was included; otherwise, 5′ was included. Spliced transcript assemblies in each cluster were used to determine whether the cluster was transcribed from the sense (S) or antisense (AS) strand of the reference gene; however, it was not determined if the cluster contained both sense and antisense transcripts or no spliced transcript. If multiple clusters were associated with the same reference gene and they could not be distinguished from the sense or antisense tag, then distance (in kb) between the clusters and the reference gene was further appended.

Differences in lincRNA expression. The expression of a lincRNA cluster was measured by the number of ‘reads’ from the associated islands normalized by the island’s size and total ‘read’ number. Clusters with different expression in two conditions were identified by EdgeR (false-discovery rate, <0.05; change in expression of over 1.5-fold in either ‘direction’ (upregulated or downregulated)44. The calculation of differences in expression required that the cluster be expressed at least in one of the two conditions; a cluster was deemed as being expressed in a sample if any of the associated islands showed enrichment for RNA-Seq ‘reads’ from the sample in both duplicates. For consistency, the same rules were applied to determine differences in expression for protein-encoding genes.

Analysis of enrichment for terms of the GO project. The online DAVID bioinformatics resource (version 6.7)45 and/or the GOrrila gene-ontology enrichment analysis and visualization tool46 was (were) used to analyze enrichment for terms of the GO project. For large-scale analysis (as in Supplementary Fig. 3c), GO term annotations were downloaded from the Mouse Genome Informatics website 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. The TopHat splice junction mapper for RNA-Seq reads47 and Cufflinks RNA-Seq analysis tools48 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 Swiss-Protein database by BlastX and the Coding Potential Calculator (CPC)24 were used to evaluate the coding potential of a transcript. If the strand information was available, only the forward three reading frames were considered, otherwise all six reading frames were considered. A transcript may be considered to encode a protein if it shows protein-level similarity to any annotated coding genes (E value < 0.0001, identity > 30%) and/or the CPC score is above zero23. However, noncoding genes are enriched for degenerate transposable elements49, we disqualified a transcript as being coding if it had any ‘hit’ with proteins associated with transposons and manually examined the CPC output to check if the determination that a transcript was coding was caused by similarity to a transposon–associated protein.

Statistics. We applied two-sided Kolmogorov-Smirnov test to assess the difference in transcription factor binding and gene expression between any two groups of lincRNAs. This test is a nonparametric method with which to test whether two probability distributions differ. It requires no prior knowledge about the distributions50. The χ2 test was used for comparison of two portions from independent samples, presented as a percentage. Comparison of means was done by a t-test without an assumption of equal variance.

Analysis of enrichment for terms of the GO project. The online DAVID bioinformatics resource (version 6.7)45 and/or the GOrrila gene-ontology enrichment analysis and visualization tool46 was (were) used to analyze enrichment for terms of the GO project. For large-scale analysis (as in Supplementary Fig. 3c), GO term annotations were downloaded from the Mouse Genome Informatics website 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. The TopHat splice junction mapper for RNA-Seq ‘reads’47 and Cufflinks RNA-Seq analysis tools48 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 Swiss-Protein database by BlastX and the Coding Potential Calculator (CPC)24 were used to evaluate the coding potential of a transcript. If the strand information was available, only the forward three reading frames were considered, otherwise all six reading frames were considered. A transcript may be considered to encode a protein if it shows protein-level similarity to any annotated coding genes (E value < 0.0001, identity > 30%) and/or the CPC score is above zero23. However, noncoding genes are enriched for degenerate transposable elements49, we disqualified a transcript as being coding if it had any ‘hit’ with proteins associated with transposons and manually examined the CPC output to check if the determination that a transcript was coding was caused by similarity to a transposon–associated protein.

Statistics. We applied two-sided Kolmogorov-Smirnov test to assess the difference in transcription factor binding and gene expression between any two groups of lincRNAs. This test is a nonparametric method with which to test whether two probability distributions differ. It requires no prior knowledge about the distributions50. The χ2 test was used for comparison of two portions from independent samples, presented as a percentage. Comparison of means was done by a t-test without an assumption of equal variance.

36. Yagi, R. et al. The transcription factor GATA3 actively represses RUNX3 protein-regulated production of interferon-γ. Immunity 32, 507–517 (2010).
37. Yamane, H., Zhu, J. & Paul, W.E. Independent roles for IL-2 and GATA-3 in stimulating naive CD4+ T cells to generate a Th2-inducing cytokine environment. J. Exp. Med. 202, 793–804 (2005).
38. Ichihara, M. et al. Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. Nucleic Acids Res. 35, e123 (2007).
39. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
40. Zang, C.Z. et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics 25, 1952–1958 (2009).
41. Mei, G. et al. Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity 30, 155–167 (2009).
42. Deaton, A.M. et al. Cell type-specific DNA methylation at intragenic CpG islands in the immune system. Genome Res. 21, 1074–1086 (2011).
43. Nakayamada, S. et al. Early Th1 cell differentiation is marked by a Tfh cell-like transition. Immunity 35, 919–931 (2011).
44. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010).
45. Huang, D., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57 (2009).
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 10, 48 (2009).
47. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111 (2009).
48. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515 (2010).
49. Kelley, D. & Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. 13, R107 (2012).
50. Hollander, M. & Wolfe, D.A. Nonparametric Statistical Methods (Wiley, 1973).