Dynamic MicroRNA Gene Transcription and Processing during T Cell Development

Francis F. Kirigin, Kenneth Lindstedt, Maclean Sellars, Maria Ciofani, Siao Li Low, Lachlan Jones, Fiona Bell, Florencia Pauli, Richard Bonneau, Richard M. Myers, Dan R. Littman and Mark M. W. Chong

*J Immunol* 2012; 188:3257-3267; Prepublished online 29 February 2012; doi: 10.4049/jimmunol.1103175

http://www.jimmunol.org/content/188/7/3257

---

**Supplementary Material**  
http://www.jimmunol.org/content/suppl/2012/02/29/jimmunol.1103175.DC1

**References**  
This article cites 47 articles, 21 of which you can access for free at: http://www.jimmunol.org/content/188/7/3257.full#ref-list-1

**Subscription**  
Information about subscribing to *The Journal of Immunology* is online at: http://jimmunol.org/subscription

**Permissions**  
Submit copyright permission requests at: http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**  
Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/alerts
Dynamic MicroRNA Gene Transcription and Processing during T Cell Development

Francis F. Kirigin,*† Kenneth Lindstedt, ‡ Maclean Sellars,* Maria Ciofani,* Siao Li Low, † Lachlan Jones, † Fiona Bell, ‡ Florencia Pauli, ‡ Richard Bonneau, †§ Richard M. Myers, †§ Dan R. Littman,*§ and Mark M. W. Chong†§

By disrupting microRNA (miRNA) biogenesis, we previously showed that this pathway is critical for the differentiation and function of T cells. Although various cloning studies have shown that many miRNAs are expressed during T cell development, and in a dynamic manner, it was unclear how comprehensive these earlier analyses were. We therefore decided to profile miRNA expression by next generation sequencing. Furthermore, we profiled miRNA expression starting from the hematopoietic stem cell. This analysis revealed that miRNA expression during T cell development is extremely dynamic, with 645 miRNAs sequenced, and the expression of some varying by as much as 3 orders of magnitude. Furthermore, changes in precursor processing led to altered mature miRNA sequences. We also analyzed the structures of the primary miRNA transcripts expressed in T cells and found that many were extremely long. The longest was pri–miR-29b-1/29a at ~168 kb. All the long pri-miRNAs also displayed extensive splicing. Our findings indicate that miRNA expression during T cell development is both a highly dynamic and a highly regulated process. The Journal of Immunology, 2012, 188: 3257–3267.

Early hematopoiesis occurs in the bone marrow (BM), giving rise to progenitors of all leukocyte lineages. Thymocyte progenitors then leave the BM to seed the thymus where definitive T cell development occurs. Early thymocytes first progress through four stages termed double-negative (DN) 1–4, because they lack expression of the CD4 and CD8 coreceptors. DN3 is a key stage. TCR β rearrangement occurs here, and productive rearrangement is required for progression to the DN4 stage (1). DN4 thymocytes then rapidly proliferate and upregulate both coreceptors as they progress to the CD4⁺CD8⁺ double-positive (DP) stage. Appropriate selection then leads to differentiation of mature CD8⁺ cytotoxic T cells or CD4⁺ T helper cells (2).

Much of what is understood about T cell development is centered on proteins, such as those involved in transcription and signal transduction. However, there is an increasing appreciation for the role of non-protein–coding RNAs (ncRNAs). Both long ncRNAs (lncRNAs) (3) and small ncRNAs (4–6) are expressed in T cells, but the function of most ncRNA classes remains unknown. The best understood class is the microRNAs (miRNAs), which are ~22-nt small RNAs that inhibit the translation of protein-coding messenger RNAs (mRNAs). Thousands of miRNAs have so far been identified in plants and animals, including some 1424 and 720 in humans and mice, respectively (miRBase) (7). miRNAs target protein-coding mRNAs via incomplete base pairings (8). Because only partial complementarity is required, numerous miRNAs can be the target of each miRNA.

miRNAs originate from long primary transcripts (known as pri-miRNAs) containing one or more secondary stem-loop structures. It is from this stem loop that the mature miRNA is eventually derived. In the canonical biogenesis pathway, the intermediate “pre-miRNA” stem loop is released from the pri-miRNA in the nucleus by the microprocessor complex. At the core of this complex is the RNase III enzyme Drosha (9). The excised pre-miRNA is then exported to the cytoplasm where it is further processed by another RNase III enzyme complex containing Dicer, which clips off the loop structure (10). This is followed by loading of the miRNA-5p:miRNA-3p duplex (i.e., the two arms of the stem loop) into the RNA-induced silencing complex (11), at which point one strand is degraded.

We and others have shown that Drosha and Dicer, and therefore miRNAs, are important throughout the T cell compartment. miRNAs are required for early thymocytes to progress through the DN stages (12). miRNAs are also necessary for T cell function. In particular, they maintain the suppressor program of Foxp3⁺ regulatory T cells. Mice with Foxp3⁺ cell-specific Drosha or Dicer
deficiency die of a lymphoproliferative multiorgan inflammatory disease because of loss of suppressor function (13–15).

Compared with the posttranscriptional processing, less is known about the transcription of miRNA genes. This has been due, in large part, to a lack of information about the full-length primary transcript of most miRNAs, and thus, genomic databases have not been able to annotate these genes. It is estimated that 30–40% of mammalian miRNAs are derived from the introns of protein-coding genes and thus are transcribed together with the host genes (16). However, the majority of miRNAs are derived from independent transcriptional units (ITUs). The transcription of most miRNA genes depends on RNA polymerase II (PolII) (17–19). Furthermore, PolII binding sites located near pre-miRNA genomic locations (potentially correlating with promoters) are marked by histone 3 trimethyl-lysine 4 (H3K4me3) (20). Because of rapid processing by the microprocessor complex, pri-miRNAs derived from these ITUs are difficult to detect in cells. Full-length primary transcripts have been characterized for only a handful of miRNAs (17, 21, 22). Thus, although it has been possible to determine from where in the genome the pre-miRNA stem-loop intermediate is derived, the structure of many miRNA genes remains undefined.

miRNAs are clearly important for the development and function of T cells. However, which specific miRNAs are important at which stage(s) is still unclear. In this article, we report the use of next generation sequencing (NGS) to construct a comprehensive miRNA atlas of T cell development. We reveal the dynamic nature of miRNA gene transcription and processing throughout this developmental pathway. We also use NGS coupled to chromatin immunoprecipitation sequencing (ChIP-seq) and polyA RNA enrichment (RNA-seq) to map the structures of miRNA genes expressed in T cells.

Materials and Methods

Mice and cell preparations

Drosophila and Dicer knockout CD4-cre mice have been described previously (15). C57BL/6 mice were purchased from Taconic Farms or produced at the Walter and Eliza Hall Institute. Mice at 6- to 8-week-old mice were used for all analyses. Single-cell suspensions were prepared from BM by flushing with a syringe, and thymocyte suspensions were prepared by mincing thymi through a 100-μm mesh. All Abs were purchased from eBioscience, and cell sorting was performed on a FACSAria (BD Biosciences).

miRNA sequencing

Total RNA was prepared from 0.5–5 × 10^6 cells (depending on population) using TRIzol (Invitrogen). The 19- to 24-nt small RNAs were fractionated, and libraries were constructed for sequencing on the Illumina GAII platform as described previously (12, 23). Sequence reads were mapped to known mature and pre-miRNAs deposited in miRBase (7). Nonmapping reads were then analyzed for potential novel miRNA species. In brief, this involved aligning the remaining reads to the Mus musculus genome (mm9 assembly, National Center for Biotechnology Information Build 37) using Novoalign (NovoCraft V2.05.04). The genomic intervals in which novel reads clustered were then extracted, and the corresponding RNA was analyzed for putative secondary structure using UNAfold (24). Sequences with predicted stem-loop structure were considered potential miRNA precursors. However, only those structures where the small RNAs actually mapped to the double-stranded stems were considered novel miRNAs. Scripts for these pipelines can be provided upon request.

RNA sequencing

RNA-seq was performed as described previously (25, 26). In brief, polyA RNAs were selected using oligo-dT magnetic beads. The RNA was fragmented and reverse transcribed to produce cDNA and then dsDNA. Illumina-compatible adaptors were ligated to the dsDNA after end-repair and dA addition. The resulting library was size selected at ~225 bp and amplified and purified before being sequenced. The reads were mapped to mouse genome (mm9 assembly) using the Illumina ELAND software. The data were then visualized using the Integrative Genomics Viewer software (27) or the University of California Santa Cruz Genome Browser (28). Mammalian sequence conservation data were also obtained from the University of California Santa Cruz Genome Browser.

ChIP sequencing

ChIPs were performed on native chromatin essentially as previously described (29), with some modifications. In brief, 1.1 × 10^7 cells were lysed in digestion buffer (50 mM Tris-HCl [pH 7.5], 250 mM sucrose, 3 mM MgCl2, 1 mM CaCl2, and 0.2% Triton X-100, supplemented with fresh 5 mM sodium butyrate, 0.5 mM PMSF, and EDTA-free protease inhibitors) at 4˚C for 10 min. The nuclei were collected by centrifugation and resuspended in fresh digestion buffer. The nuclei were warmed to 37˚C, and then, micrococcal nuclease was added to a final concentration of 0.09 U/μl. After 5 min, the reaction was stopped by the addition of EDTA/EGTA. The nuclei were collected and then lysed by two rounds of sonication (Diagenode Bioruptor 300) and incubation in lysis buffer (10 mM Tris-HCl [pH 7.5], 0.25 mM EDTA, supplemented with fresh 5 mM sodium butyrate and 0.5 mM PMSF and EDTA-free protease inhibitors) for 1 h. More than 90% mononucleosomes release was confirmed. A total of 25 μg chromatin (~5 × 10^6 cell equivalents) was subjected to ChIP using 15 μl Protein A Dynabeads (Invitrogen) precoated with anti-H3K4me3 (05-745; Upstate Biotechnology/Millipore) or anti-H3K36me3 (Ab9050; Abcam) Abs. Immunoprecipitated mononucleosomes were washed twice in radioimmunoprecipitation assay buffer, twice in radioimmunoprecipitation assay buffer plus 0.3 mM NaCl, twice in LiCl buffer (0.25 M LiCl, 1% Nonidet P-40, and 0.5% sodium deoxycholate), and finally twice with Tris-EDTA buffer. DNA was eluted by incubation with 3% SDS and 50 μg/ml proteinase K in Tris-EDTA at 65˚C for 1 h. After phenol/chloroform extraction, libraries were prepared from 10 ng DNA using the ChIP-Seq Prepreparation Kit (Illumina), according to the manufacturer’s protocol, and sequenced on the Illumina GAII platform. The reads were mapped to mouse genome (mm9 assembly) using the Illumina ELAND software.

Table I. miRNA sequencing statistics

<table>
<thead>
<tr>
<th>Population</th>
<th>Known miRNA Mappers</th>
<th>Other Genic Mappers (Unique)</th>
<th>Other Intergenic Mappers (Unique</th>
<th>Repeat Mappers</th>
<th>Total Mappers</th>
<th>miRNA as % of Unique Mappers</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSK</td>
<td>2021255</td>
<td>13741</td>
<td>8724</td>
<td>17022</td>
<td>2060742</td>
<td>98.9</td>
</tr>
<tr>
<td>MPP</td>
<td>217589</td>
<td>17760</td>
<td>10294</td>
<td>20833</td>
<td>2224782</td>
<td>98.7</td>
</tr>
<tr>
<td>LMPP</td>
<td>1186047</td>
<td>14377</td>
<td>22247</td>
<td>61127</td>
<td>1283798</td>
<td>97.0</td>
</tr>
<tr>
<td>Pro-B</td>
<td>2987281</td>
<td>12643</td>
<td>14269</td>
<td>24758</td>
<td>3038951</td>
<td>99.1</td>
</tr>
<tr>
<td>DN1</td>
<td>4202313</td>
<td>41069</td>
<td>21996</td>
<td>45273</td>
<td>4310651</td>
<td>98.5</td>
</tr>
<tr>
<td>DN2</td>
<td>2816672</td>
<td>18451</td>
<td>15482</td>
<td>36155</td>
<td>2886760</td>
<td>98.8</td>
</tr>
<tr>
<td>DN3</td>
<td>5162882</td>
<td>106602</td>
<td>85511</td>
<td>280951</td>
<td>5639396</td>
<td>96.4</td>
</tr>
<tr>
<td>DN4</td>
<td>791050</td>
<td>12478</td>
<td>15793</td>
<td>37500</td>
<td>856821</td>
<td>98.2</td>
</tr>
<tr>
<td>DP</td>
<td>623483</td>
<td>31004</td>
<td>31644</td>
<td>63082</td>
<td>749213</td>
<td>90.9</td>
</tr>
<tr>
<td>CD4SP</td>
<td>5094220</td>
<td>33593</td>
<td>19174</td>
<td>45919</td>
<td>5192906</td>
<td>99.0</td>
</tr>
<tr>
<td>CD8SP</td>
<td>3013905</td>
<td>17187</td>
<td>10608</td>
<td>24096</td>
<td>3065796</td>
<td>99.1</td>
</tr>
<tr>
<td>MEF</td>
<td>332787</td>
<td>7094</td>
<td>11602</td>
<td>17750</td>
<td>369233</td>
<td>94.7</td>
</tr>
<tr>
<td>Total</td>
<td>30407790</td>
<td>325999</td>
<td>267344</td>
<td>674466</td>
<td>31675599</td>
<td>—</td>
</tr>
<tr>
<td>Mean</td>
<td>2533983</td>
<td>27167</td>
<td>22279</td>
<td>56206</td>
<td>2639633</td>
<td>97.3</td>
</tr>
</tbody>
</table>

*LMPP, lymphoid-primed multipotent progenitor.*
RT-PCR

cDNA was obtained by reverse transcription of total RNA using Super- 
script III (Invitrogen) and random hexamers. The following primers were 
employed to amplify pri-miRNA sequences: 5'-GATACAAAAAGTCAG 
GACAGCC-3' and 5'-AGTCTCAGCCTGCCAG-3' were used to am-

plify pri-mir-21; 5'-GACTACCCTCTGCCCTTGTCG-3' and 5'-TTCTGC 
TTCTGTAGTCTGTA-3' were used to amplify the pri-let-7c-2/let-7b; 
and 5'-AGGTTTATCCTGTGGTGTGA-3' and 5'-CATACCGCAATG 
ACAAAGG-3' were used to amplify the pri–mir-29b-1/29a.

NGS data sets

All miRNA-seq, RNA-seq, and ChIP-seq data sets have been deposited with 
the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under 
accession number GSE30584.

Results

miRNA libraries and the identification of putative novel miRNAs expressed during T cell development

Early studies characterizing the expression of miRNAs in T cells 
used cloning or hybridization microarrays on whole lymphoid 
organs or partially fractionated T cell populations (5, 30). More 
recently, Neilson et al. (6) extended this analysis by focusing 
specifically on T cell development. They analyzed libraries of 
short RNA clones derived from FACS-sorted thymocyte subsets. 
However, given that <1000 clones were obtained for some pop-

ulations, it was unclear how comprehensive this analysis was. 
We therefore decided to construct a more comprehensive miRNA

FIGURE 1. Dynamic miRNA expression during T cell development. (A) The indicated populations corresponding to distinct stages of T cell de-

velopment from LSKs through to mature CD4SP and CD8SP thymocytes were FACS purified and then analyzed for miRNA expression by miRNA-seq. As a comparison, pro-B cells (related population) and MEFs (unrelated population) were also analyzed. Shown is a heat map of all mature miRNA species 
sequenced at least once. The data have been clustered based on Euclidean distance. Comparisons of miRNA profiles between CD4SP and CD8SP thy-

mocytes (B); DN1 and pro-B (C); and LSK and MEF (D). Examples of miRNAs expressed at constant levels (E); hematopoietic progenitor-specific miRNAs 
(F); miRNAs that are transiently upregulated during T cell development (G); and miRNAs with bimodal expression patterns (H). r, Pearson’s correlation.
profile of T cell development using NGS. Furthermore, we decided to start from the hematopoietic stem cell-enrich Lin$^{-}$Sca1$^{+}$ kit$^{+}$ (LSK) population in the BM. We sorted the following populations corresponding to known steps in T cell development from the BM and thymus of C57BL/6 mice: LSK, multipotent progenitor (MPP; Lin$^{-}$Sca1$^{+}$CD135$^{+}$CD106$^{-}$), lymphoid-primed multipotent progenitor (Lin$^{-}$Sca1$^{+}$CD135$^{+}$CD106$^{-}$), DN1 (CD90$^{+}$CD4$^{+}$CD8$^{-}$CD3$^{-}$CD44$^{+}$CD25$^{+}$), DN2 (CD90$^{+}$CD4$^{-}$CD8$^{-}$CD3$^{-}$CD44$^{+}$CD25$^{-}$), DN3 (CD90$^{+}$CD4$^{+}$CD8$^{+}$CD3$^{-}$/lo CD44$^{+}$CD25$^{+}$), DN4 (CD90$^{+}$CD4$^{-}$CD8$^{-}$CD3$^{-}$/lo CD44$^{+}$CD25$^{+}$), DP (CD90$^{+}$CD4$^{+}$CD8$^{+}$CD3$^{-}$/lo), CD4SP (CD90$^{+}$CD4$^{-}$CD8$^{+}$CD24$^{+}$/lo CD3$^{+}$/lo), and CD8SP (CD90$^{+}$CD4$^{+}$CD8$^{+}$CD24$^{+}$/lo CD3$^{+}$/lo). We also included pro-B cells (B220$^{+}$IgM$^{-}$CD43$^{+}$/hi) as a related population and murine embryonic fibroblasts (MEFs) as an unrelated population. A total of $3.17 \times 10^7$ mappable reads was obtained at an average of $2.64 \times 10^6$ reads per library (Table I). The vast majority (90.0–99.1%) mapped to previously annotated miRNA loci (Table I). An average of 2.1% of reads were non-unique mappers (i.e., mapped to multiple locations within the genome, such as repetitive sequences) and were thus excluded from further analyses. An additional 1.9% of reads were found to be unique mappers, equally distributed between genic and intergenic regions.

To identify novel miRNAs, the remaining 1.9% of unique reads were pooled and analyzed for groups of reads clustering within 50- to 200-bp genomic intervals. The RNA sequence of these intervals was then analyzed for potential secondary structure in unafold (24). As pre-miRNAs are expected to fold into stable secondary stem-loop structures, we analyzed further only those sequences with predicted free energies ($\Delta G$) of less than or equal to $-30$ kcal/mol. This yielded 310 sequences with potentially stable secondary structure, of which 66 displayed clear stem-loop

**FIGURE 2.** Altered miRNA precursor processing at DN3 to DP stages of T cell development. (A) Expression levels of miR-106b, -93, and -25, which are all derived from the same polycistron. Distribution of mature miRNA subspecies derived from the pre–mir-17 (B) and pre–mir-21 (C) precursors. The data indicate the contribution each subspecies makes to the total mature miR-17 or -21 pool at each stage of T cell development. miRNAs derived from both 5p and 3p arms are shown.
The Journal of Immunology

structure and mammalian sequence conservation (data not shown). Despite so many sequences predicted to fold into pre-miRNA-like stem-loops, for only 20 did the reads map clearly to the stems (Supplemental Fig. 1). Three of these (numbers 77, 1123, and 1413) have since been annotated as miR-1843b, miR-3068 and miR-1306b by miRBase (7). Two (numbers 28 and 665) mapped to predicted small nucleolar RNAs (31), and five (numbers 199, 624, 647, 671, and 823) mapped to predicted transfer RNAs (32). However, these previous annotations were based on genomic sequence rather than the RNA. Analysis of the actual RNAs suggests that all fold into stem-loop structures typical of pre-miRNAs rather than small nucleolar RNAs or transfer RNAs (Supplemental Fig. 1).

The dynamics of miRNA expression during T cell development

miRNA-seq generated a comprehensive and quantitative atlas of miRNA expression during T cell development (Fig. 1A, Supplemental Table I). A total of 645 different mature miRNAs (including miRNA* species) were sequenced at least once in one of the populations within the T cell developmental pathway (Supplemental Table I). Previous studies suggested that although most miRNAs are widely expressed (4), different sets of miRNAs are expressed by different cell types (33). These studies also implied that similar miRNA profiles are shared by related cell types. From our miRNA-seq analysis, developmental relationships were clearly evident. The highly related populations, CD4SP and CD8SP thymocytes, displayed highly similar miRNA profiles, with a Pearson’s correlation of 0.954 (Fig. 1B). The closely related populations, DN1 thymocytes and pro-B cells, both early being stages in their respective developmental pathways, displayed somewhat similar miRNA profiles, with a Pearson’s correlation of 0.75 (Fig. 1C). In contrast, the completely unrelated populations, LSK and MEFs, displayed completely different miRNA profiles, with a Person’s correlation of 0.04 (Fig. 1D).

Although many miRNAs were expressed at a constant level throughout (Fig. 1E), including miR-181d and many Let-7 members, others displayed highly dynamic patterns. Many of the miRNAs that were highly expressed in the LSK and MPP stages, including miR-10a and miR-126-3p, were downregulated by as much as 3 orders of magnitude by the CD4SP and CD8SP thymocytes (Fig. 2A). However, miR-181a expression was not surprising because miR-181a/b/c expression was not a surprise. According to miRBase (7), miR-181c clusters with miR-181d and thus coordinate expression might have been expected. However, because miR-181d displayed a constant expression pattern (Fig. 1E), there appeared to be a discrepancy between the two linked miRNAs. Finally, numerous miRNAs displayed biphasic expression patterns, including miR-142-5p and miR-142-3p to be expressed with similar patterns as both are derived from the same pre-miRNA. These miR-142-5p/3p and miR-181d discords suggest that miRNA expression might not simply be determined by transcription of the gene but also on precursor processing that appears to vary during T cell development.

Developmental stage-specific processing of miRNA precursors

To further explore the possibility of regulated miRNA precursor processing, we examined the expression of miR-106b, miR-93, and miR-25, which are all derived from intron 13 of the Mcm7 host transcript. Because all three are derived from the same precursor, one might expect similar expression patterns if all are processed with similar efficiencies. miR-106b was poorly expressed at all stages (Fig. 2A). miR-93 and -25 were expressed at similar levels at most stages (Fig. 2A). However, miR-25 was downregulated at the DN3 to DP stages, suggesting that there was a loss in the ability to process the premir-25 intermediate specifically at these stages.

Fine-mapping of miRNA subspecies also suggested altered precursor processing of some miRNAs at the DN3 to DP stages. The 19- to 22-mer AAAGCUUACAGUGCGAGGU(AGU) subspecies were the dominant miR-17-5p molecules expressed in LSK to DN2, CD4SP, and CD8SP populations (Fig. 2B). However, in DN3 thymocytes, there was a shift of 2 bases 5’ to the 21- to 24-mer CAAAGCUUACAGUGCGAGGU(AGU) miR-17-5p subspecies and the appearance of miR-17-3p subspecies. This shift by 1 base again 3’ in DN4 and DP thymocytes to 20- to 22-mer AAAAGCUUACAGUGCGAGGU(AGU) miR-17-5p subspecies.

![FIGURE 3](http://www.jimmunol.org/) Impact of altered seed sequence on predicted targets. (A) Shown are the miR-17 and miR-21 subspecies sequenced. The seed sequences are highlighted. (B) The miR-17+1 subspecies has the same seed sequence as miR-302a and miR-106a, whereas the miR-17+2 subspecies has the same seed sequence as miR-302c. Venn diagrams of targets predicted by Targetscan. Indicated are the numbers of targets that predicted for the annotated seed of miR-17 (C) or miR-21 (D) and the +1 and +2 shifted seed. Overlaps indicate identical targets predicted for more than one subspecies.
Pre–miR-21 also displayed altered processing at the DN3 to DP stages (Fig. 2C). Although most stages expressed 19- to 21-mer GUUAUCAGACUGAUGUUG(AC) miR-21-5p subspecies, this was shifted by 2 base positions 5’ to 21- to 23-mer UAGCUUA-UCAGACUGAUGUUG(AC) in DN3 thymocytes and by 1 base in DN4 and DP thymocytes to 18- to 22-mer AGCUUAUCAGA-

CUGAUGU(UGAC). Subspecies derived from the 3p arm again appeared in DN4 and DP thymocytes.

Thus, not only is miRNA maturation regulated during T cell development but also the precise cut sites. Interestingly, for both maturation of the polycistron mir-106b∼25 and variable cleavage sites, the shifts occurred specifically at the DN3 to DP stages,
suggested that there may be specific changes to the miRNA biogenesis machinery at this point in the T cell development.

**Impact of altered miRNA precursor processing on the predicted targets**

Although miRNAs pair imperfectly with target mRNAs, pairing between the seed sequence (nucleotide positions 2–8) of the miRNA and target is thought to be a critical determinant (8). Altered precursor processing that result in shifts at the 5'-seed sequence may therefore affect target recognition.

Upon more careful examination of the miR-17 subspecies expressed during T cell development, we found that those expressed only in DN3 thymocytes corresponded to the annotated mature miR-17 in miRBase (7). The miR-17 subspecies expressed in DN4 and DP thymocytes were shifted 3' by 1 base, whereas all other stages expressed miR-17 subspecies that were shifted 3' by 2 bases compared with the annotated miR-17 (Fig. 3A). A +1 shift results in a seed sequence identical to that of miR-302a or miR-106a, whereas a shift of +2 results in a seed sequence identical to that of miR-302c (Fig. 3B). Targetscan (8) predicts 827 targets for miR-17, whereas 498 targets are predicted for a +1 shift (Fig. 3C, Supplemental Table II). Of the targets predicted for miR-17+1, only 286 are also predicted targets of the annotated miR-17. In other words, a single nucleotide change in the seed sequence of miR-17 changed two-thirds of the predicted targets. With a +2 shift, only 132 of the same targets were predicted.

For miR-21, the impact of shifted seeds on target prediction was even more dramatic (Fig. 3D, Supplemental Table II). As with miR-17, the miR-21 subspecies expressed in DN3 thymocytes corresponded to the mature miR-21 in miRBase (7). Only 7 of the 94 predicted targets of the +1 shifted miR-21 were the same as for the annotated miR-21. Similarly, only 7 of the 63 predicted targets of the +2 shifted miR-21 were the same. Although only computational predictions, these data suggest that altered precursor processing could potentially have an enormous impact on the targets that are regulated by any single miRNA. Further genetic studies will be required to determine whether altered mature miRNA sequence does indeed affect target recognition this dramatically.

**The structures of miRNA genes**

The dynamic miRNA expression patterns observed during T cell development also suggested regulation at the level of miRNA gene transcription. In contrast to posttranscriptional processing, less is known about the transcription of miRNAs. Although up to 40% of miRNAs may be derived from the introns of protein-coding genes (16), and thus dependent on the transcription of the host gene, most are derived from ITUs. From the few pri-miRNA transcripts cloned previously from ITUs, we know that transcription is dependent on PolII and that they contain 7-methyl guanosine caps and polyA tails (17, 21). However, for most ITU-derived miRNAs, the pri-miRNA transcripts have yet to be cloned. This is thought to be attributable to rapid processing of these transcripts by the miRNA biogenesis machinery (17).

To map the structures of pri-miRNAs and miRNA genes in T cells, we first needed to block miRNA processing and allow transcript accumulation. We and others (13, 17, 34) have previously shown that knocking out/down components of the microprocessor complex results in the accumulation of pri-miRNA transcripts that can be detected, at least, by Northern blotting. Deletion of Drosha in T cells results in the accumulation of pri-miRNA transcripts derived from ITUs that can also be detected by RNA-seq (Fig. 4A, 4B) or by quantitative RT-PCR (data not shown). Drosha deficiency did not appear to affect the expression of mRNAs with intron-embedded miRNAs (data not shown). RNA-seq analysis of polyA RNAs in Drosha-deficient cells thus allowed us to visualize the pri-miRNA transcripts and the structure of actively transcribed miRNA genes. We analyzed libraries from both CD4+ and CD8+ T cells, and ~2 x 10⁷ reads were obtained for each library.

To clearly delineate the miRNA genes, we also analyzed specific histone modifications. Barski et al. (35) have previously shown that miRNA genes carry chromatin modifications similar to that of protein-coding genes. Mapping of H3K4me3 peaks can therefore be used to identify the promoters of actively transcribed miRNA genes, whereas broadly distributed H3K36me3 should mark the gene body (Fig. 4C–F). From the miRNA-seq profile of CD4SP and CD8SP cells, we identified some 70 miRNA genomic loci (Supplemental Table III). Of these, 35 were clearly located within introns of protein-coding miRNAs, 24 loci were clearly expressed as ITUs, and an additional 9 loci appeared to map within or close to exons of protein-coding genes.

Several of the 24 miRNA ITUs corresponded to previously cloned pri-miRNAs (22, 36), including mir-17~92a (Fig. 4B). However, to our knowledge, for the majority miRNAs, this was...
the first demonstration of the primary transcript. Although several miRNA genes were only a few kilobases in length, such as Mir-142 (Fig. 4C) and Mir-146a (Supplemental Table III), most were substantially larger (Fig. 4D, 4E, Supplemental Table III). The largest, included Mir-29b-1/29a, which spanned 167.8kb, and Mir-181a-2/181b-2, which spanned 111.4kb (Supplemental Table III). Furthermore, several miRNA genes displayed more than one H3K4me3 peak (Fig. 4E), suggesting that more than one promoter may be used by these genes.

Splicing of pri-miRNA transcripts has been reported previously (21). We found that although short pri-miRNA transcripts, such as mir-142, were not spliced, long pri-miRNAs appeared to be spliced extensively (Fig. 4D, 4E). RT-PCR across putative introns confirmed splicing throughout the flanking ssRNA strands (Fig. 5A, 5B).

It was curious to observe so much splicing. Extensive splicing is reminiscent of IncRNA maturation (37), in addition to protein-coding mRNAs. Thus, we wondered whether these long ssRNAs that flank pre-miRNAs might remain in cells as mature IncRNAs. Our RNA-seq analysis had been performed on polyA-enriched RNA. In wild-type cells, processing would cleave off these flanking ssRNAs and thus would not have been picked up the oligo-dT pulldown and subsequent RNA-seq. We therefore measured expression by quantitative RT-PCR. However, we found that these flanking ssRNAs did not accumulate in either wild-type

**FIGURE 6.** Identification of independent pri-miRNA transcriptional units embedded within protein-coding genes. (A) The Mir-21 gene overlaps with 3'UTR of the overlapping protein-coding Vmp1 gene. The pri–mir-21 transcript can be identified as an independent transcript by the presence of an additional H3K4me3 peak (marking a promoter) within intron 10 and RNA transcription through introns 10 and 11 of the host Vmp1 gene. Additional examples of miRNA genes overlapping with protein-coding genes include Mir-700 (B), Mir-15a/16-1 (C), and Mir-1891 (D). (E) Confirmation that the pri–mir-21 transcript is transcribed independently from the overlapping protein-coding gene. Shown are the locations of the primers used to detect the pri–mir-21 transcript. One primer is located within intron 10 of Vmp1, and thus specific for pri–mir-21. (F) The pri–mir-21 transcript accumulates only in Drosha-deficient CD4+ T cells, indicating that it is a bona fide pri-miRNA transcript. All analyses were performed on CD4+ T cells. (–), Gene on the minus strand.
or Dicer-deficient cells, suggesting that they are normally degraded following Drosha-dependent cleavage (Fig. 5C).

**Independent pri-miRNA transcriptional units embedded within protein-coding genes**

Nine loci appeared to map within or close to exons of other genes (Supplemental Table III). miR-21 is annotated within the 3′-UTR (exon 12) of Vmp1 in mice (Fig. 6A). However, the presence of an H3K4me3 peak within intron 10 and RNA transcription through introns 10 and 11 of the Vmp1 gene suggested that miR-21 is actually derived from a short ITU that overlaps with the 3′-end of Vmp1. Indeed, the same overlapping structure was reported for MIR21 and VMP1 in humans (21). Such an overlap between independently transcribed miRNA and protein-coding genes was not unique. We found eight other examples (Supplemental Table III), including Mir-700, Mir-15a/16-1, and Mir-1839 (Fig. 6B–D).

By using primers located within the pri-miR-21 sequence but not exons of the overlapping Vmp1 gene (Fig. 6E), we could detect the accumulation of the transcript in Drosha-deficient T cells (Fig. 6F), thus confirming that it is indeed a bona fide pri-miRNA transcript. Furthermore, the accumulation of RNA from intron 10 of Vmp1 specifically within Drosha-deficient cells allowed us to confirmed that the H3K4me3 peak within intron 10 is most likely the Mir-21 promoter.

**miRNA genes display poor evolutionary sequence conservation**

We next analyzed the sequence conservation of the miRNA genes expressed in T cells to better understand the structures of these genes. In contrast to protein-coding genes, miRNA genes were found to display very poor evolutionary conservation (Fig. 7A, 7B). Within the transcript itself, only the pre-miRNA portions displayed a high degree of conservation among mammals, whereas the remainder of the transcript displayed little or no conservation. This indicates that it is only the sequence of the stem loop under evolutionary pressure to maintain sequence. This is likely due to the requirement for base pairings between the 5p and 3p arms to change specifically at late DN and DP thymocyte stages. Precise endonucleolytic cleavage sites were also altered. Changes in the expression of mature miRNAs derived from the same transcribed polycistron suggested altered precursor processing. This appeared to be due to method of detection. Li et al. (40) reported expression at early DN stages and not in DP thymocytes. In contrast, Neilson et al. (6) found highest expression in DN4 thymocytes but overall greatest enrichment in DP thymocytes. The reason for this discrepancy is unclear but may be due to method of detection. Li et al. analyzed miR-181a expression by PCR, whereas we and the Neilson study used cloning. Regardless, miR-181a is thought to regulate the expression of numerous proteins important for TCR signaling and thymocyte selection, including phosphatases (40) and CD69 (6). It has also been shown that miR-181a is necessary for the negative selection of autoreactive TCRs (41), confirming its expression and function in DP thymocytes.

Although most of the changes in miRNA profiles can be attributed to likely changes in pri-miRNA transcription, discrepancies in the expression of mature miRNAs derived from the same polycistron suggested altered precursor processing. This appeared to change specifically at late DN and DP thymocyte stages. Likewise, precise endonucleolytic cleavage sites were also altered.

**Discussion**

**miRNA dynamics and T cell development**

miRNA-seq revealed substantially more dynamic miRNA expression during T cell development than previously found. Several striking patterns were observed. First, many of the miRNAs that were highly expressed in BM progenitor populations, such as miR-10a, miR-99h, and so on, were downregulated by 2–3 orders of magnitude in thymocyte populations. A survey of miRNA-seq data of B cell development and various other mature hematopoietic lineages (38) indicate that these miRNAs are also downregulated in other differentiation pathways. These miRNAs may therefore play important functions in maintaining the multipotential state of hematopoietic progenitors. miR-125a has been shown previously to amplify the size and activity of hematopoietic stem cells in mice, in part, by inhibiting the expression of the proapoptotic protein Bak (39). We found that all miR-125 family members, miR-125a, miR-125b, and miR-351, were highly expressed in hematopoietic progenitors but then downregulated. Because all three have the same seed sequence, it is likely that they have similar spectra of targets. Coordinate expression in progenitors then downregulation was also observed for miR-221 and miR-222, which share the same seed sequence. miR-99a and miR-99b also displayed this same expression pattern. Simultaneous expression of miRNAs with shared seed sequence may function as a mechanism of redundancy in hematopoietic progenitors.

Changes in miRNA expression profiles specifically at late DN and DP thymocyte stages were also prominent. For example, miR-181a, miR-181b, and miR-181c were all transiently upregulated. High levels of miR-181a have been reported previously for the thymus. Li et al. (40) reported expression at early DN stages and not in DP thymocytes. In contrast, Neilson et al. (6) found highest expression in DN4 thymocytes but overall greatest enrichment in DP thymocytes. The reason for this discrepancy is unclear but may be due to method of detection. Li et al. analyzed miR-181a expression by PCR, whereas we and the Neilson study used cloning. Regardless, miR-181a is thought to regulate the expression of numerous proteins important for TCR signaling and thymocyte selection, including phosphatases (40) and CD69 (6). It has also been shown that miR-181a is necessary for the negative selection of autoreactive TCRs (41), confirming its expression and function in DP thymocytes.

**FIGURE 7.** Poor evolutionary conservation of miRNA genes. Shown is the alignment of H3K4me3 (marking the promoter) and RNA with mammalian sequence conservation for two miRNA genes. Analyses were performed on Drosha-deficient CD4+ T cells. The miRNA gene domains (pri-miRNA) are indicated by the arrows. Note that only the regions corresponding to the mature miRNAs, the promoter, and a few other interspersed regions display sequence conversion.
It has previously been observed that miRNAs are usually expressed with fixed 5’-ends but more flexible 3’-ends (42, 43). The interpretation was that a flexible 5’-end would alter seed sequence and therefore target genes. We found that a single homogenous population of cells at the same stage in T cell development invariably expressed miRNAs with identical 5’-ends. However, the precise ends could change in DN3, DN4, and/or DP thymocytes. We found that even a single nucleotide shift in the seed sequence dramatically affected the predicted targets (at least those predicted by Targetscan). Future genetic studies will be required to determine the impact of such altered precursor processing on mature miRNA targets. Regardless, our data imply that simply determining whether a miRNA is expressed or not is insufficient for mining whether a miRNA is expressed or not is insufficient for determining the specificity of that miRNA are known.

That both polycistron processing and 5’-ends were altered at the same stages in T cell development suggests that the miRNA biogenesis machinery may be regulated differently in this study. Interestingly, these changes appear to correlate with the phenotypes observed when Drosha or Dicer was deleted specifically within the T cell developmental pathway. Lack of either enzyme, and therefore miRNAs, resulted in a developmental block at the DN3 stage (12). In the case of miR-17 and miR-21, their 5’-ends are generated by Drosha-mediated cleavage. Similarly, the release of different pre-miRNA intermediates from a common polycistron is dependent on Drosha. The Drosha complex may differ within thymocytes at different stages of development. Cell-specific factors could be recruited. Indeed, p68 and p72 DEAD-box RNA helicases have been shown to bind to Drosha and modulate the processing of specific miRNAs (44). Thus, miRNA expression is a complex regulated process and not simply dependent on transcription and inexcisable processing by Drosha and Dicer.

miRNA gene structure

Previous studies have attempted to map pri-miRNA transcripts by assembling expressed sequence tags (45) or transcriptional features surrounding miRNA genomic locations, to predict the putative boundaries (46). However, because most primary transcripts are efficiently processed by the microprocessor complex, mapping unobservable transcripts was obviously difficult. Although Saini et al. (46) predicted that some miRNA genes would indeed be very large, we had expected most pri-miRNAs to be only several hundred base pairs to a few kilobases in length, as the few characterized primary miRNAs were within this range (17, 21). Pri-miRNA prediction based on EST assembly also suggested this (45). This was already surprisingly long, given that premiRNA intermediates are 50–70 nt and only 10 nt of ssRNA immediately 5’ and 3’ of the stem loop are necessary for processing, at least in vitro (47). To our surprise, we found that most of the miRNA genes expressed within T cells were tens of kilobases and up to a 167-kb length. Given that, on average, each miRNA gene encodes two pre-miRNA stem loops, only ~88 bp (5p + 3p of each stem loop) of each transcript is potentially transcribed. This means that <99.95% of a pri-miRNA is transcribed, processed and then discarded.

It is unclear why pri-miRNAs transcripts are so long. One possibility is that the remaining long ssRNAs function as IncRNAs. However, the poor sequence conservation of these transcripts and their degradation in Drosha-sufficient cells suggest that this is unlikely. Another possibility is that flanking ssRNA strands are required for regulated pri-miRNA processing. cis elements might function from a distance to regulate the recognition and processing of the stem loop in an analogous manner to distal transcription factor binding sites that regulate gene transcription. At least in vitro, stem-loop-containing RNAs are not processed by Drosha at equivalent efficiencies (12, 47). Although the majority of these flanking ssRNA strands have poor sequence conservation, peaks of conservation were present throughout. These peaks may correspond to RNA cis elements critical for regulated processing. These elements might function in a cell-specific manner and could help explain the variations in miRNA processing that occur during T cell development.

Our study has shown that miRNA expression during T cell development is highly dynamic. This dynamic nature is likely to be fundamental for controlling the transcriptional and proteomic landscape during T cell development. It has long been recognized that transcription factors can both activate and repress gene expression. However, with the discovery of regulatory ncRNAs, such as miRNAs, it is clear that these transcriptional networks are more complicated than previously appreciated. Transcription factors that activate the transcription of protein-coding genes may, in some instances, suppress targets indirectly by activating the transcription of regulatory miRNAs. Likewise, repressive transcription factors may upregulate protein expression by repressing miRNAs. Thus, a precise understanding both of protein-coding gene regulation and of miRNA (and other ncRNA) gene regulation will be fundamental to elucidating the molecular mechanisms that control the development and function of T cells.

Disclosures

The authors have no financial conflicts of interest.

References


