

Published in final edited form as:

*J Immunol.* 2012 April 1; 188(7): 3257–3267. doi:10.4049/jimmunol.1103175.

## Dynamic microRNA gene transcription and processing during T cell development

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

<sup>\*</sup>The Kimmel Center for Biology and Medicine of the Skirball Institute, New York University School of Medicine, New York, NY 10016, USA

<sup>†</sup>New York University Center for Genomics and Systems Biology, New York, NY 10003, USA

<sup>‡</sup>The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC 3052, Australia

<sup>§</sup>HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA

<sup>¶</sup>Courant Institute for Mathematical Sciences, New York, NY 10003, USA

<sup>||</sup>Howard Hughes Medical Institute, New York University School of Medicine, New York, NY 10016, USA

<sup>#</sup>Department of Medical Biology, University of Melbourne, Parkville, VIC 3010, Australia

### Abstract

By disrupting microRNA (miRNA) biogenesis, we previously showed that this pathway is critical for the differentiation and function of T cells. While 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 means of 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 ~168kb. 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.

### INTRODUCTION

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 to 4, because they lack expression of the CD4 and CD8 coreceptors. DN3 is a key stage. T cell receptor (TCR)  $\beta$  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<sup>+</sup>CD8<sup>+</sup>

double positive (DP) stage. Appropriate selection then leads to differentiation of mature CD8<sup>+</sup> cytotoxic T cells or CD4<sup>+</sup> helper T 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 ~22nt 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 1,424 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 mRNAs 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 from. 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 (RISC) (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<sup>+</sup> regulatory T cells. Mice with FoxP3<sup>+</sup> cell-specific Drosha or Dicer deficiency die from a lymphoproliferative multi-organ inflammatory disease due to loss of suppressor function (13-15).

Compared to the post-transcriptional 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. Here, 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 (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

*Drossha*<sup>F/F</sup> and *Dicer*<sup>F/F</sup> *CD4-cre* mice have been previously described (13). 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 100µm mesh. All antibodies were purchased from eBioscience and cell sorting was performed on a FACS-Aria (Beckton Dickinson).

### MiRNA Sequencing

Total RNA was prepared from  $0.5\text{--}5\times 10^5$  cells (depending on population) using Trizol (Invitrogen). The 19-24nt small RNAs were fractionated, and libraries were constructed for sequencing on the Illumina GAII platform as previously described (12, 23). Sequence reads were mapped to known mature and pre-miRNAs deposited in miRBase (7). Non-mapping reads were then analyzed for potential novel miRNA species. In brief, this involved aligning the remaining reads to the *Mus musculus* genome (mm9 assembly, NCBI 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 previously described (25, 26). In brief, polyA RNAs were selected using oligo-dT magnetic beads. The RNA was fragmented and reversed transcribed to produce cDNA, then dsDNA. Illumina-compatible adaptors were ligated to the dsDNA after end-repair and dA addition. The resulting library was size selected at ~225bp, amplified and purified before being sequenced. The reads were mapped to mouse genome (mm9 assembly) using the Illumina ELAND software. The data was then visualized using the Integrative Genomics Viewer software (27) or UCSC Genome Browser (28). Mammalian sequence conservation data was also obtained from the UCSC Genome Browser.

### ChIP sequencing

ChIPs were performed on native chromatin essentially as described (29), with some modifications. In brief,  $1.1\times 10^7$  cells were lysed in digestion buffer (50mM Tris-HCl pH 7.5, 250mM sucrose, 3mM MgCl<sub>2</sub>, 1mM CaCl<sub>2</sub>, and 0.2% Triton-X100, supplemented with fresh 5mM sodium butyrate, 0.5mM PMSF and EDTA-free Protease inhibitors) at 4°C for 10min. The nuclei were collected by centrifugation and resuspended in fresh digestion buffer. The nuclei were warmed to 37°C, then micrococcal nuclease was added to a final concentration of 0.09U/mL. After 5min, the reaction was stopped by the addition of EDTA/EGTA. The nuclei were collected, then lysed by 2 rounds of sonication (Diagenode Bioruptor 300) and incubation in lysis buffer (10mM Tris-HCl pH 7.5, 0.25mM EDTA, supplemented with fresh 5mM sodium butyrate, 0.5mM PMSF and EDTA-free Protease inhibitors) for 1h. >90% mononucleosomes release was confirmed. 25µg chromatin (~ $5\times 10^6$  cell equivalents) was subjected to ChIP using 15µL Protein A Dynabeads (Invitrogen) pre-coated with anti-H3K4me3 (Upstate/Millipore, 05-745) or anti-H3K36me3 (Abcam, Ab9050) antibodies. Immunoprecipitated mononucleosomes were washed twice in RIPA buffer, twice in RIPA buffer plus 0.3M NaCl, twice in LiCl buffer (0.25 M LiCl, 1% NP40,

0.5% sodium deoxycholate), and finally twice with TE buffer. DNA was eluted by incubation with 3% SDS and 50 µg/mL proteinase K in TE at 65°C for 1h. After phenol/chloroform extraction, libraries were prepared from 10ng 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.

## RT-PCR

cDNA was obtained by reverse transcription of total RNA using Superscript III (Invitrogen) and random hexamers. The following primers were employed to amplify pri-miRNA sequences: 5'-GATACCAAAATGTCAGACAGCC and 5'-AGTCTCAGCTGCACTAGC were used to amplify pri-mir-21; 5'-GACTACGCTCTGCCCTTCTG and 5'-TTCTGCTTCCTGAGTGCTGA were used to amplify the pri-mir-29b-1/29a; 5'-AGGTTTCATACCTGTTGGTGGA and 5'-CATTACGGCAATGACCAAGG were used to amplify the pri-let-7c-2/let-7b.

## Next Generation Sequencing Data sets

All miRNA-seq, RNA-seq and ChIP-seq datasets have been deposited with the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession #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 and co-workers extended this analysis by focusing specifically on T cell development (6). They analyzed libraries of short RNA clones derived from FACS sorted thymocyte subsets. However, given that fewer than 1000 clones were obtained for some populations, it was unclear how comprehensive this analysis was. We therefore decided to construct a more comprehensive miRNA profile of T cell development using NGS. Furthermore, we decided to start from the hematopoietic stem cell (HSC)-enrich Lin<sup>-</sup>Sca1<sup>+</sup>kit<sup>+</sup> (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<sup>-</sup>Sca1<sup>+</sup>kit<sup>+</sup>CD135<sup>+</sup>CD106<sup>+</sup>), lymphoid-primed multipotent progenitor (LMPP, Lin<sup>-</sup>Sca1<sup>+</sup>kit<sup>+</sup>CD135<sup>+</sup>CD106<sup>-</sup>), DN1 (CD90<sup>+</sup>CD4<sup>-</sup>CD8<sup>-</sup>CD3<sup>-</sup>CD44<sup>+</sup>CD25<sup>-</sup>), DN2 (CD90<sup>+</sup>CD4<sup>-</sup>CD8<sup>-</sup>CD3<sup>-</sup>CD44<sup>+</sup>CD25<sup>+</sup>), DN3 (CD90<sup>+</sup>CD4<sup>-</sup>CD8<sup>-</sup>CD3<sup>-</sup>CD44<sup>-</sup>CD25<sup>+</sup>), DN4 (CD90<sup>+</sup>CD4<sup>-</sup>CD8<sup>-</sup>CD3<sup>lo</sup>CD44<sup>-</sup>CD25<sup>-</sup>), DP (CD90<sup>+</sup>CD4<sup>+</sup>CD8<sup>+</sup>CD3<sup>lo</sup>), CD4SP (CD90<sup>+</sup>CD4<sup>+</sup>CD8<sup>-</sup>CD24<sup>lo</sup>CD3<sup>hi</sup>) and CD8SP (CD90<sup>+</sup>CD4<sup>-</sup>CD8<sup>+</sup>CD24<sup>lo</sup>CD3<sup>hi</sup>). We also included pro-B cells (B220<sup>+</sup>IgM<sup>-</sup>CD43<sup>hi</sup>) as a related population, and murine embryonic fibroblasts (MEFs) as an unrelated population. A total of 3.17×10<sup>7</sup> mappable reads was obtained, at an average of 2.64×10<sup>6</sup> reads per library (Table 1). The vast majority (90.0-99.1%) mapped to previously annotated miRNA loci (Table 1). 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. A further 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 50-200bp genomic intervals. The RNA sequence of these intervals were 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  $-30\text{kcal/mol}$ . This yielded 310 sequences with potentially stable secondary structure, of which 66 displayed clear stem-loop structure and mammalian sequence conservation (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 (Figure S1). Three of these (#77, #1123 and #1413) have since been annotated as miR-1843b, miR-3068 and miR-1306b by miRBase (7). Two (#28 and #665) mapped to predicted snoRNAs (31) and 5 (#199, #624, #647, #671 and #823) mapped to predicted tRNAs (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 snoRNAs or tRNAs (Figure S1).

### The dynamics of miRNA expression during T cell development

MiRNA-seq generated a comprehensive and quantitative atlas of miRNA expression during T cell development (Figure 1A and Table S1). 645 different mature miRNAs (including miRNA\* species) were sequenced at least once in one of the populations within the T cell developmental pathway (Table S1). Previous studies suggested that while 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 (Figure 1B). The closely related populations, DN1 thymocytes and Pro-B cells, both being early steps in their respective developmental pathways, displayed somewhat similar miRNA profiles, with a Pearson's correlation of 0.75 (Figure 1C). In contrast, the completely unrelated populations, LSK and MEFs, displayed completely different miRNA profiles, with a Pearson's correlation of 0.04 (Figure 1D).

While many miRNAs were expressed at a constant level throughout (Figure 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 stages (Figure 1F). Others were transiently upregulated at specific stages. In particular, there was a noticeable spike in miR-181a, miR-181b, miR-181c and miR-142-3p at the late DN to DP stages (Figure 1G). It has previously been observed that miR-181 family members are highly enriched in DP thymocytes (6). Thus, the spike in the miR-181a/b/c expression was not a surprise. According to miRBase (7), miR-181c clusters with miR-181d and thus co-ordinate expression might have been expected. However, since miR-181d displayed a constant expression pattern (Figure 1E), there appeared to be a discrepancy between the two linked miRNAs. Finally, numerous miRNAs displayed biphasic expression patterns, including miR-142-5p (Figure 1H). Again, one would have expected 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-181c/d discordances 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. Since 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 (Figure 2A). MiR-93 and 25 were expressed at similar



levels at most stages (Figure 2A). However, miR-25 was downregulated at the DN3 to DP stages, suggesting that there was a loss in the ability to process the pre-miR-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 19-22-mer AAGUGCUUACAGUGCAGGU(AGU) subspecies were the dominant miR-17-5p molecules expressed in LSK to DN2, CD4SP and CD8SP populations (Figure 2B). However, in DN3 thymocytes there was a shift of 2 bases 5' to the 21-24-mer CAAAGUGCUUACAGUGCAGGU(AGU) miR-17-5p subspecies, and the appearance of miR-17-3p subspecies. This shifted by 1 base again 3' in DN4 and DP thymocytes to 20-22-mer AAAGUGCUUACAGUGCAGGU(AGU) miR-17-5p subspecies.

Pre-miR-21 also displayed altered processing at the DN3 to DP stages (Figure 2C). While most stages expressed 19-21-mer GCUUAUCAGACUGAUGUUG(AC) miR-21-5p subspecies, this was shifted by 2 base positions 5' to 21-23-mer UAGCUUAUCAGACUGAUGUUG(AC) in DN3 thymocytes, and by 1 base in DN4 and DP thymocytes to 18-22-mer AGCUUAUCAGACUGAUGU(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, suggesting 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 to 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, while all other stages expressed miR-17 subspecies that were shifted 3' by 2 bases compared to the annotated miR-17 (Figure 3A). A +1 shift results in a seed sequence identical to that of miR-302a or miR-106a, while a shift of +2 results in a seed sequence identical to that of miR-302c (Figure 3B). Targetscan (8) predicts 827 targets for miR-17, while 498 targets are predicted for a +1 shift (Figure 3C and Table S2). 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 (Figure 3D and Table S2). Like for 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 if 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 post-transcriptional 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 PolIII 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 due to rapid processing of these transcripts by the miRNA biogenesis machinery<sup>17</sup>.

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 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 (13, 17, 34). 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 (Figures 4A and B) 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<sup>+</sup> and CD8<sup>+</sup> T cells, and  $\sim 2 \times 10^7$  reads were obtained for each library.

To clearly delineate the miRNA genes, we also analyzed specific histone modifications. Barski and co-workers have previously shown that miRNA genes carry chromatin modifications similar to that of protein-coding genes (35). Mapping of H3K4me3 peaks can therefore be employed to identify the promoters of actively transcribed miRNA genes, while broadly distributed H3K36me3 should mark the gene body (Figures 4C to F). From the miRNA-seq profile of CD4SP and CD8SP cells, we identified some 70 miRNA genomic loci (Table S3). Of these, 35 were clearly located within introns of protein-coding mRNAs, 24 loci were clearly expressed as ITUs, and a further 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 (Figure 4B). However, for the majority miRNAs, this was the first demonstration of the primary transcript. While several miRNA genes were only a few kilobases in length, such as *Mir-142* (Figure 4C) and *Mir-146a* (Table S3), most were substantially larger (Figure 4D and E, and Table S3). The largest, included *Mir-29b-1/29a*, which spanned 167.8kb, and *Mir-181a-2/181b-2*, which spanned 111.4kb (Table S3). Furthermore, several miRNA genes displayed more than one H3K4me3 peak (Figure 4E), suggesting that more than one promoter may be employed by these genes.

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

It was curious to observe so much splicing. Extensive splicing is reminiscent of long ncRNA 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 lncRNAs. Our RNA-seq analysis had been performed on polyA-enriched RNA. In wild-type cells, processing would cleave off these flanking ssRNA strands 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 wildtype or Dicer deficient cells, suggesting that they are normally degraded following Drosha-dependent cleavage (Figure 5C).

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

Nine loci appeared to map within or close to exons of other genes (Table S3). MiR-21 is annotated within the 3'UTR (exon 12) of *Vmp1* in mice (Figure 6A). However, the presence of a 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 *VMPI* in humans (21). Such an overlap between independently transcribed miRNA and protein-coding genes was not unique. We found 8 other examples (Table S3), including *Mir-700*, *Mir-15a/16-1* and *Mir-1839* (Figure 6B to D).

By using primers located within the pri-mir-21 sequence but not exons of the overlapping *Vmp1* gene (Figure 6E), we could detect the accumulation of the transcript in Drosha deficient T cells (Figure 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 confirm 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 (Figure 7A and B). Within the transcript itself, only the pre-miRNA portions displayed a high degree of conservation among mammals, while 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 form the stem-loop structure and be processed by Drosha and Dicer, and to pair with target protein-coding mRNAs.

In contrast to the transcripts themselves, the regions upstream of miRNA genes displayed substantial sequence conservation. Regions immediately 5' of transcriptional start sites, corresponding to the promoters of the genes, were invariably conserved (Figure 7A and B). We also observed numerous additional upstream peaks of conservation that are likely to correspond to *cis* regulatory elements. Thus, only the pre-miRNA stem-loop and *cis* regulatory elements of miRNA genes are evolutionarily conserved and the majority of the RNA itself permits a large degree of sequence flexibility.

## 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. Firstly, many of the miRNAs that were highly expressed in bone marrow progenitor populations, such as miR-10a, miR-99b, etc, were downregulated by 2 to 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 previously



been shown to amplify the size and activity of HSCs in mice, in part, by inhibiting the expression of the pro-apoptotic 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 previously been reported for the thymus. Li and co-workers reported expression at early DN stages and not in DP thymocytes (40). In contrast, Neilson and co-workers found highest expression in DN4 thymocytes, but overall greatest enrichment in DP thymocytes (6). The reason for this discrepancy is unclear, but may be due to method of detection. Li and co-workers analyzed miR-181a expression by PCR, whereas we and the Neilson study employed 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.

While 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. 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 implies that simply determining whether a miRNA is expressed or not is insufficient for investigating miRNA-dependent gene regulation. It is also essential that the precise sequence(s) 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 here. 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 stage of development. Cell-specific co-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 inexorable processing by Drosha and Dicer.

## MiRNA gene structure

Previous studies have attempted to map pri-miRNA transcripts by assembling ESTs (45) or transcriptional features surrounding miRNA genomic locations, in order 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 and co-workers predicted that some miRNA genes would indeed be very large (46), we had expected most pri-miRNAs to be only several hundred base pairs to a few kilobases in length, as the few characterized pri-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 pre-miRNA intermediates are 50-70nt and only 10nt 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 167kb length. Given that, on average, each miRNA gene encodes two pre-miRNA stem-loops, only ~88bp (5p + 3p of each stem-loop) of each transcript is potentially functional. This means that up 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 lncRNAs. However, the poor sequence conservation of these transcripts and their degradation in Drosha-sufficient cells suggests that this is unlikely. Another possibility is that flanking RNA 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.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## REFERENCES

1. Rothenberg EV, Moore JE, Yui MA. Launching the T-cell-lineage developmental programme. *Nat Rev Immunol.* 2008; 8:9–21.
2. Klein L, Hinterberger M, Wirnsberger G, Kyewski B. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nat Rev Immunol.* 2009; 9:833–844.

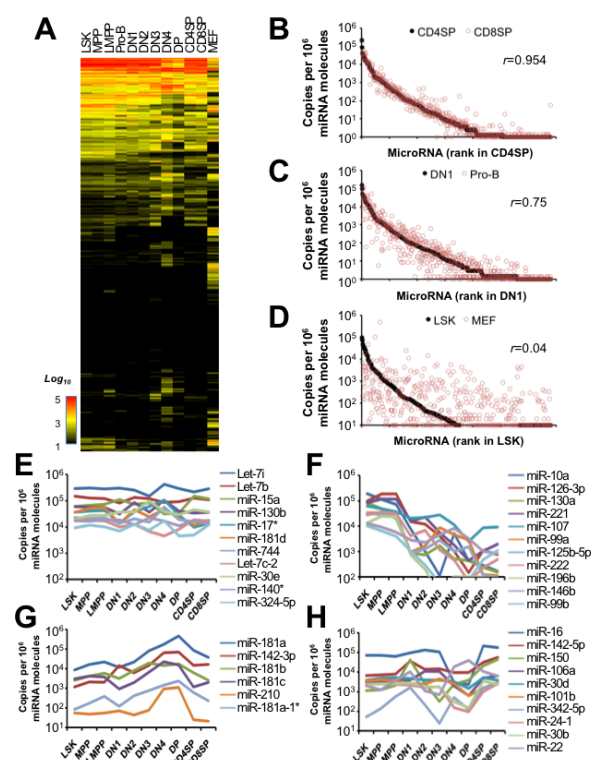
3. Pang KC, Dinger ME, Mercer TR, Malquori L, Grimmond SM, Chen W, Mattick JS. Genome-wide identification of long noncoding RNAs in CD8+ T cells. *J Immunol.* 2009; 182:7738–7748.
4. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foa R, Schliwka J, Fuchs U, Novosel A, Muller RU, Schermer B, Bissels U, Inman J, Phan Q, Chien M, Weir DB, Choksi R, De Vita G, Frezzetti D, Trompeter HI, Hornung V, Teng G, Hartmann G, Palkovits M, Di Lauro R, Wernet P, Macino G, Rogler CE, Nagle JW, Ju J, Papavasiliou FN, Benzing T, Lichter P, Tam W, Brownstein MJ, Bosio A, Borkhardt A, Russo JJ, Sander C, Zavolan M, Tuschl T. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell.* 2007; 129:1401–1414.
5. Monticelli S, Ansel KM, Xiao C, Socci ND, Krichevsky AM, Thai TH, Rajewsky N, Marks DS, Sander C, Rajewsky K, Rao A, Kosik KS. MicroRNA profiling of the murine hematopoietic system. *Genome Biol.* 2005; 6:R71.
6. Neilson JR, Zheng GX, Burge CB, Sharp PA. Dynamic regulation of miRNA expression in ordered stages of cellular development. *Genes Dev.* 2007; 21:578–589.
7. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 2011; 39:D152–157.
8. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell.* 2005; 120:15–20.
9. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN. The nuclear RNase III Drosha initiates microRNA processing. *Nature.* 2003; 425:415–419.
10. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science.* 2001; 293:834–838.
11. Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell.* 2005; 123:607–620.
12. Chong MM, Zhang G, Cheloufi S, Neubert TA, Hannon GJ, Littman DR. Canonical and alternate functions of the microRNA biogenesis machinery. *Genes Dev.* 2010; 24:1951–1960.
13. Chong MM, Rasmussen JP, Rudensky AY, Littman DR. The RNaseIII enzyme Drosha is critical in T cells for preventing lethal inflammatory disease. *J Exp Med.* 2008; 205:2005–2017.
14. Liston A, Lu LF, O'Carroll D, Tarakhovsky A, Rudensky AY. Dicer-dependent microRNA pathway safeguards regulatory T cell function. *J Exp Med.* 2008; 205:1993–2004.
15. Zhou X, Jeker LT, Fife BT, Zhu S, Anderson MS, McManus MT, Bluestone JA. Selective miRNA disruption in T reg cells leads to uncontrolled autoimmunity. *J Exp Med.* 2008; 205:1983–1991.
16. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. *Genome Res.* 2004; 14:1902–1910.
17. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* 2004; 23:4051–4060.
18. Corcoran DL, Pandit KV, Gordon B, Bhattacharjee A, Kaminski N, Benos PV. Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data. *PLoS One.* 2009; 4:e5279.
19. Wang G, Wang Y, Shen C, Huang YW, Huang K, Huang TH, Nephew KP, Li L, Liu Y. RNA polymerase II binding patterns reveal genomic regions involved in microRNA gene regulation. *PLoS One.* 2010; 5:e13798.
20. Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG, Zhang X, Song JS, Fisher DE. Chromatin structure analyses identify miRNA promoters. *Genes Dev.* 2008; 22:3172–3183.
21. Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA.* 2004; 10:1957–1966.
22. Eis PS, Tam W, Sun L, Chadburn A, Li Z, Gomez MF, Lund E, Dahlberg JE. Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proc Natl Acad Sci U S A.* 2005; 102:3627–3632.
23. Hafner M, Landgraf P, Ludwig J, Rice A, Ojo T, Lin C, Holoch D, Lim C, Tuschl T. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods.* 2008; 44:3–12.

24. Markham NR, Zuker M. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol.* 2008; 453:3–31.
25. Reddy TE, Pauli F, Sprouse RO, Neff NF, Newberry KM, Garabedian MJ, Myers RM. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res.* 2009; 19:2163–2171.
26. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008; 5:621–628.
27. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol.* 2011; 29:24–26.
28. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res.* 2002; 12:996–1006.
29. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell.* 2007; 129:823–837.
30. Chen CZ, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. *Science.* 2004; 303:83–86.
31. Pruitt KD, Tatusova T, Klimke W, Maglott DR. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.* 2009; 37:D32–36.
32. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997; 25:955–964.
33. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of tissue-specific microRNAs from mouse. *Curr Biol.* 2002; 12:735–739.
34. Wang Y, Medvid R, Melton C, Jaenisch R, Blelloch R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nature genetics.* 2007; 39:380–385.
35. Barski A, Jothi R, Cuddapah S, Cui K, Roh TY, Schones DE, Zhao K. Chromatin poises miRNA- and protein-coding genes for expression. *Genome Res.* 2009; 19:1742–1751.
36. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM. A microRNA polycistron as a potential human oncogene. *Nature.* 2005; 435:828–833.
37. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature.* 2009; 458:223–227.
38. Kuchen S, Resch W, Yamane A, Kuo N, Li Z, Chakraborty T, Wei L, Laurence A, Yasuda T, Peng S, Hu-Li J, Lu K, Dubois W, Kitamura Y, Charles N, Sun HW, Muljo S, Schwartzberg PL, Paul WE, O'Shea J, Rajewsky K, Casellas R. Regulation of microRNA expression and abundance during lymphopoiesis. *Immunity.* 2010; 32:828–839.
39. Guo S, Lu J, Schlanger R, Zhang H, Wang JY, Fox MC, Purton LE, Fleming HH, Cobb B, Merkenschlager M, Golub TR, Scadden DT. MicroRNA miR-125a controls hematopoietic stem cell number. *Proc Natl Acad Sci U S A.* 2010; 107:14229–14234.
40. Li QJ, Chau J, Ebert PJ, Sylvester G, Min H, Liu G, Braich R, Manoharan M, Soutschek J, Skare P, Klein LO, Davis MM, Chen CZ. miR-181a is an intrinsic modulator of T cell sensitivity and selection. *Cell.* 2007; 129:147–161. [PubMed: 17382377]
41. Ebert PJ, Jiang S, Xie J, Li QJ, Davis MM. An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a. *Nat Immunol.* 2009; 10:1162–1169.
42. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, Blelloch R, Schroth GP, Nusbaum C, Bartel DP. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.* 2010; 24:992–1009.
43. Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell.* 2006; 127:1193–1207.
44. Fukuda T, Yamagata K, Fujiyama S, Matsumoto T, Koshida I, Yoshimura K, Mihara M, Naitou M, Endoh H, Nakamura T, Akimoto C, Yamamoto Y, Katagiri T, Foulds C, Takezawa S, Kitagawa H, Takeyama K, O'Malley BW, Kato S. DEAD-box RNA helicase subunits of the

Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat Cell Biol.* 2007; 9:604–611.

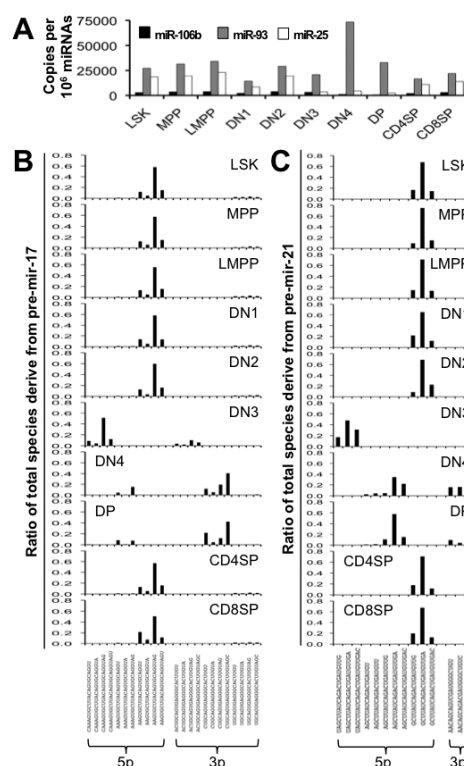
45. Gu J, He T, Pei Y, Li F, Wang X, Zhang J, Zhang X, Li Y. Primary transcripts and expressions of mammal intergenic microRNAs detected by mapping ESTs to their flanking sequences. *Mamm Genome.* 2006; 17:1033–1041.
46. Saini HK, Enright AJ, Griffiths-Jones S. Annotation of mammalian primary microRNAs. *BMC Genomics.* 2008; 9:564.
47. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell.* 2006; 125:887–901.



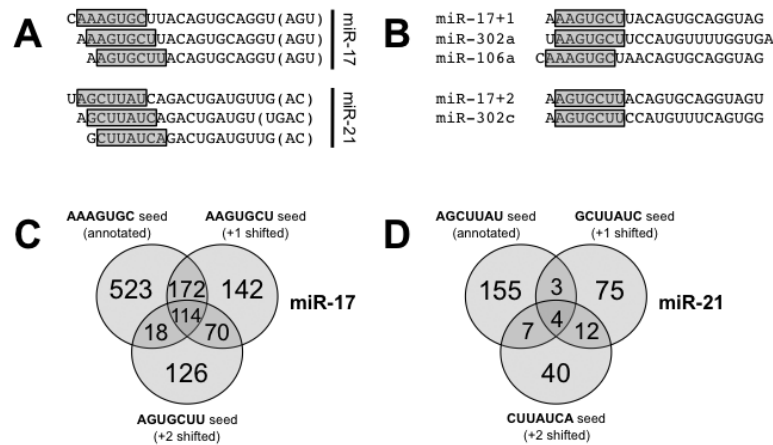


**Figure 1. Dynamic miRNA expression during T cell development**

(A) The indicated populations corresponding to distinct stages of T cell development from LSKs through to mature CD4SP and CD8SP thymocytes were FACS-purified 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 heatmap of all mature miRNA species sequenced at least once. The data has been clustered based on Euclidean distance. Comparisons of miRNA profiles between (B) CD4SP and CD8SP thymocytes; (C) DN1 and Pro-B; and (D) LSK and MEF.  $r$  = Pearson's correlation. Examples of (E) miRNAs expressed at constant levels; (F) hematopoietic progenitor-specific miRNAs; (G) miRNAs that are transiently upregulated during T cell development; and (H) miRNAs with bimodal expression patterns.

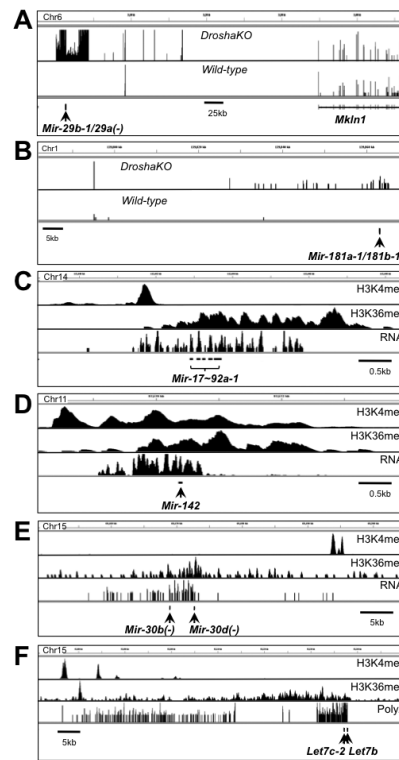


**Figure 2. Altered miRNA precursor processing at DN3 to DP stages of T cell development**  
**(A)** Expression levels of miR-106b, 93 and 25, which are all derived from the same polycistron. Distribution of mature miRNA subspecies derived from the **(B)** pre-mir-17 and **(C)** pre-mir-21 precursors. The data indicates 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.



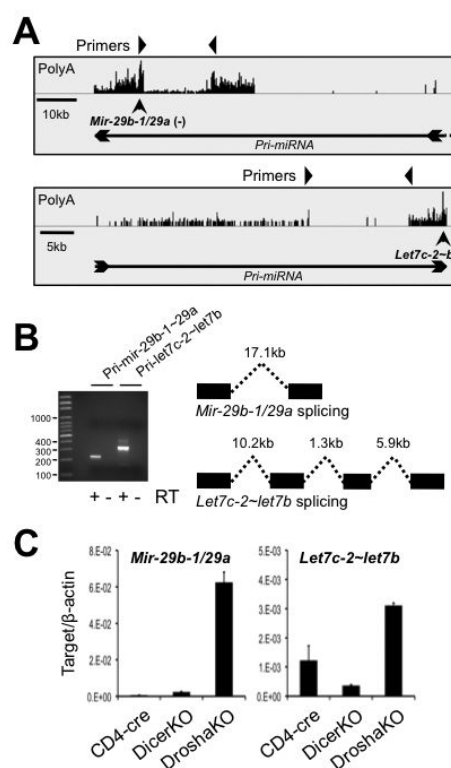
**Figure 3. Impact of altered seed sequence on predicted targets**

(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 (C) miR-17 or (D) miR-21, and the +1 and +2 shifted seed. Overlaps indicate identical targets predicted for more than one subspecies.



**Figure 4. Mapping the structure of miRNA genes expressed in T cells**

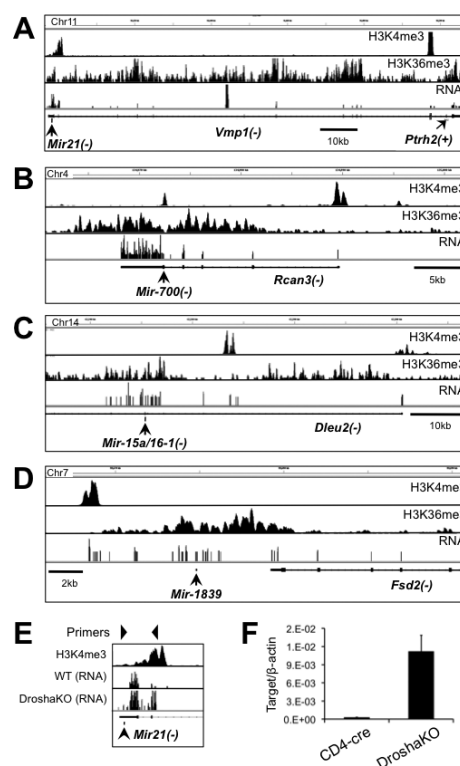
Drosha deficiency results in the accumulation of pri-miRNA transcripts derived from independent transcriptional units. Shown are the bicistronic (A) *Mir-29b-1/29a* and (B) *Mir-181a-1/181b-1* genes in CD4<sup>+</sup> T cells. PolyA RNAs (i.e. full-length and spliced RNAs) were isolated from MACS-purified CD4<sup>+</sup> T cells obtained from *Drosha*<sup>F/F</sup> *CD4-cre* and wildtype mice. Libraries were then constructed and analyzed by NGS (RNA-seq). The RNA reads were then mapped to the mouse genome. (–) = Gene on the minus strand. Shown is the relative RNA expression in WT and Drosha deficient CD4<sup>+</sup> T cells. Note that the neighboring protein-coding *Mkln1* gene is not affected by Drosha deficiency. (B) RNA-seq combined with ChIP-seq for histone modification facilitates the mapping of actively transcribed miRNA genes. ChIPs for the histone marks H3K4me3 and H3K36me3 were performed on nucleosomes isolated from CD4<sup>+</sup> T cells. Libraries were then constructed and analyzed by NGS (ChIPseq). The RNA or ChIP reads were then mapped to the mouse genome. A H3K4me3 peak marks a promoter, while H3K36me3 is broadly deposited over a gene body. The boundary of a transcriptional unit is defined by the interval in which both the pri-miRNA transcript (RNAs-seq) and H3K36me3 map to, and the presence of a H3K4me3 peak at the 5' end. Shown are examples of small miRNA genes (C) *Mir-17-92a-1* and (D) *Mir-142*; and large miRNA genes (E) *Mir-30d/30b* and (F) *Let7c-2/let7b*.



**Figure 5. Extensive splicing of pri-miRNA transcripts**

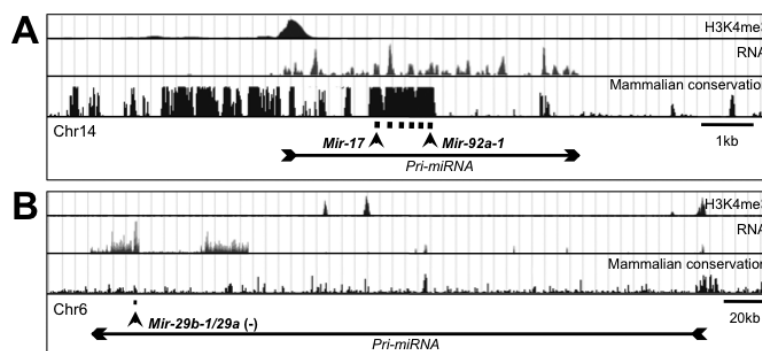
(A) Location of intron-spanning primer pairs within the *Mir-29b-1/29a* and *Let7c-2/let7b* genes. (B) RT-PCR was performed on RNA extracted from Drosha deficient CD4<sup>+</sup> T cells. PCR products were cloned and sequenced to determine splice junctions. (C) Flanking RNA strands were quantified by quantitative RT-PCR in Drosha deficient, Dicer deficient and control CD4<sup>+</sup> T cells.





**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: (B) *Mir-700*, (C) *Mir-15a/16-1*, and (D) *Mir-1891*. (–) = Gene on the minus strand. (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<sup>+</sup> T cells, indicating that it is a *bona fide* pri-miRNA transcript. All analyses were performed on CD4<sup>+</sup> T cells.



**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<sup>+</sup> 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 conservation.

**Table 1**  
**MicroRNA sequencing statistics**

| <i>Population</i> | <i>Known<br/>miRNA<br/>mappers</i> | <i>Other genic<br/>mappers<br/>(unique)</i> | <i>Other<br/>Intergenic<br/>mappers<br/>(unique)</i> | <i>Repeat<br/>mappers</i> | <i>Total mappers</i> | <i>MicroRNA as<br/>% of unique<br/>mappers</i> |
|-------------------|------------------------------------|---------------------------------------------|------------------------------------------------------|---------------------------|----------------------|------------------------------------------------|
| <b>LSK</b>        | 2021255                            | 13741                                       | 8724                                                 | 17022                     | 2060742              | 98.9                                           |
| <b>MPP</b>        | 217589                             | 17760                                       | 10294                                                | 20833                     | 2224782              | 98.7                                           |
| <b>LMPP</b>       | 1186047                            | 14377                                       | 22247                                                | 61127                     | 1283798              | 97.0                                           |
| <b>Pro-B</b>      | 2987281                            | 12643                                       | 14269                                                | 24758                     | 3038951              | 99.1                                           |
| <b>DN1</b>        | 4202313                            | 41069                                       | 21996                                                | 45273                     | 4310651              | 98.5                                           |
| <b>DN2</b>        | 2816672                            | 18451                                       | 15482                                                | 36155                     | 2886760              | 98.8                                           |
| <b>DN3</b>        | 5162882                            | 106602                                      | 85511                                                | 280951                    | 5635946              | 96.4                                           |
| <b>DN4</b>        | 791050                             | 12478                                       | 15793                                                | 37500                     | 856821               | 96.5                                           |
| <b>DP</b>         | 623483                             | 31004                                       | 31644                                                | 63082                     | 749213               | 90.9                                           |
| <b>CD4SP</b>      | 5094220                            | 33593                                       | 19174                                                | 45919                     | 5192906              | 99.0                                           |
| <b>CD8SP</b>      | 3013905                            | 17187                                       | 10608                                                | 24096                     | 3065796              | 99.1                                           |
| <b>MEF</b>        | 332787                             | 7094                                        | 11602                                                | 17750                     | 369233               | 94.7                                           |
| <b>Total</b>      | <b>30407790</b>                    | <b>325999</b>                               | <b>267344</b>                                        | <b>674466</b>             | <b>31675599</b>      | <b>-</b>                                       |
| <b>Mean</b>       | <b>2533983</b>                     | <b>27167</b>                                | <b>22279</b>                                         | <b>56206</b>              | <b>2639633</b>       | <b>97.3</b>                                    |