Analysis of Transcription Factor Expression during Discrete Stages of Postnatal Thymocyte Differentiation

Sahba Tabrizifard, Alexandru Olaru, Jason Plotkin, Mohammad Fallahi-Sichani, Ferenc Livak and Howard T. Petrie

*J Immunol* 2004; 173:1094-1102; doi: 10.4049/jimmunol.173.2.1094

http://www.jimmunol.org/content/173/2/1094

Supplementary Material http://www.jimmunol.org/content/suppl/2004/07/02/173.2.1094.DC1

Why *The JI*?

- **Rapid Reviews! 30 days** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Speedy Publication!** 4 weeks from acceptance to publication

*average

References This article cites 64 articles, 34 of which you can access for free at: http://www.jimmunol.org/content/173/2/1094.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

The Journal of Immunology is published twice each month by The American Association of Immunologists, Inc., 1451 Rockville Pike, Suite 650, Rockville, MD 20852 Copyright © 2004 by The American Association of Immunologists All rights reserved. Print ISSN: 0022-1767 Online ISSN: 1550-6606.
Analysis of Transcription Factor Expression during Discrete Stages of Postnatal Thymocyte Differentiation

Sahba Tabrizifard,* Alexandru Olaru,† Jason Plotkin,* Mohammad Fallahi-Sichani,* Ferenc Livak,‡‡ and Howard T. Petrie‡*,§

Postnatal T lymphocyte differentiation in the thymus is a multistage process involving serial waves of lineage specification, proliferative expansion, and survival/cell death decisions. Although these are believed to originate from signals derived from various thymic stromal cells, the ultimate consequence of these signals is to induce the transcriptional changes that are definitive of each step. To help to characterize this process, high density microarrays were used to analyze transcription factor gene expression in RNA derived from progenitors at each stage of T lymphopoietic differentiation, and the results were validated by a number of appropriate methods. We find a large number of transcription factors to be expressed in developing T lymphocytes, including many with known roles in the control of differentiation, proliferation, or cell survival/death decisions in other cell types. Some of these are expressed throughout the developmental process, whereas others change substantially at specific developmental transitions. The latter are particularly interesting, because stage-specific changes make it increasingly likely that the corresponding transcription factors may be involved in stage-specific processes. Overall, the data presented here represent a large resource for gene discovery and for confirmation of results obtained through other methods. The Journal of Immunology, 2004, 173: 1094–1102.

The continuous, lifelong production of new T cells by the thymus is an essential requirement in the maintenance of immune function and homeostasis (1). Thymocyte differentiation is a stepwise process initiated by bone marrow progenitors that home to the thymus via the blood. The earliest intrathymic progenitors are defined by the absence of mature T lineage markers (CD4CD8 double-negative; DN3). This population is heterogeneous and can be divided into multiple stages (reviewed in Ref. 2).

The earliest of these, designated DN stage 1 (DN1), can be identified by the expression of CD44 and CD117 and by the absence of CD25. DN1 cells are not T lineage restricted and can give rise to multiple other lineages under appropriate conditions (3–6). After an average of ~10 days of intrathymic residence (7, 8), DN1 cells asynchronously leave the perimedullary regions, where they first enter the thymus (8, 9), and enter the cortex proper (8). This movement corresponds to differentiation into the DN2 stage and is marked by the acquisition of CD25. Cells at the DN2 stage are more lineage restricted, although they can still give rise to non-T lineages (3–6). In contrast, cells at the next stage of development, designated DN3 (CD2544low117low), bear the irreversible hallmark of T lineage commitment, in the form of rearrangements of the TCR loci (10, 11), particularly DJβ rearrangements (11).

Nonetheless, the thymus generates numerous T lineage cell types, including CD4, CD8, γδ, NK-T, and T-regulatory cell types (to name a few), and it remains unclear at what point these lineages diverge. Thus, although DN3 cells are T lineage restricted, additional lineage specification events continue to be required even after this stage.

Differentiation onward from the DN3 stage appears to occur only in cells that have generated productive rearrangements at one or more of the TCR loci (12, 13) and coincides with expression of both CD4 and CD8 mature lineage markers (double-positive; DP).

The early part of the DP phase (preDP, commonly but inappropriately referred to as DN4) is characterized by cells with low surface expression of CD4 and CD8 (14), as well as by the presence of complete, productive Vβ-DJβ rearrangements (12). The TCRβ protein is also expressed on the cell surface at this time (15), in conjunction with an invariant pre-T α-chain (16) and CD3 (14). Rearrangement of the TCRα locus is then initiated (11), and CD4 and CD8 accumulate at high levels on the cell surface. Cyto genesis then ceases for the most part (17, 18), resulting in the production of nondividing, small DP cells (smDP). Further differentiation is associated with selection of those cells from this pool that express TCRs of the correct specificity. By this point, each thymic homing progenitor has undergone ~1 million-fold expansion in cell number (7), representing 20 serial divisions. This proliferation is not synchronized throughout the differentiation process. The largest number of cell divisions occurs at the DN1 stage (7), consistent with the relatively long cell cycle times (19, 20) and extended period of development (8) associated with this stage. DN2 and DN3 cells both divide a few times each (7) but cycle at very different rates (19, 20), indicating further differences in their genetic programs. The preDP stage is associated with almost as many cell divisions as DN1 cells (7), but in contrast to DN1, cell cycle times appear to be extremely short (19–21), given that these cells must expand 6- to 8-fold in <2 days before withdrawal from cycle (7).
The numerous complex changes that accompany lymphopoiesis in the postnatal thymus are initiated by signals from stromal cells that constitute the thymic microenvironment (reviewed in Ref. 22). Ultimately, however, cellular differentiation is defined by changes in the expression patterns of the required genes. Gene expression can be controlled in a variety of ways, but a major regulatory mechanism is the modulation of transcription, mediated by nuclear transcription factors. Transcription factors bind to specific cis-regulatory DNA elements (promoters, enhancers, and silencers) to induce or repress gene expression (23). Transcription factors present within a cell act in regulatory networks to establish various patterns of gene expression, and changes in these patterns define cellular differentiation (24). Consequently, the identification of transcriptional regulatory networks present in precursor cells vs their progeny may be used to reveal the mechanisms that drive differentiation. Recent studies have examined gene expression during hematopoiesis (25, 26) but have not focused on specific subsets of genes in specific developmental stages. To this end, we have performed gene expression analyses in progenitor thymocyte stages associated with the lineage commitment and proliferation processes, using high density microarrays. Here we summarize the results with regard to transcription factors that are expressed at different progenitor stages, with particular emphasis on those that change with respect to various developmental transitions. This analysis reveals novel patterns of expression of numerous transcription factors, including many not previously known to be involved in the thymocyte differentiation process. We discuss some examples of these with respect to their known roles in other cell types and potential corresponding function in thymocyte progenitors, and/or their relationships to upstream or downstream components known to be required by developing thymocytes. In addition, the data provided here constitute a rich informational resource for interrogation and validation of transcriptional patterns in developing lymphocytes.

Materials and Methods

Cell sorting and RNA extraction

Thymocytes were isolated from 4- to 5-wk-old male C57BL/6 mice (The Jackson Laboratory, Bar Harbor, ME). Early lymphopoietic progenitors were isolated by first depleting small thymocytes (DP cells) using an iso-osmotic gradient of 13.4% Opti-Prep (Greiner Bio-One, Longwood, FL) in HBSS, followed by staining with a lineage mixture of Abs (CD3, CD4, CD8, CD11b, Gr1, TER-119) and depletion with paramagnetic beads. Populations were identified as follows: DN1, CD24<sup>+</sup>CD25<sup>-</sup>44<sup>+</sup>; DN2, CD24<sup>+</sup>CD25<sup>-</sup>44<sup>-</sup>; DN3, CD24<sup>+</sup>CD25<sup>-</sup>44<sup>+</sup>; preDP, CD24<sup>+</sup>CD25<sup>-</sup>44<sup>+</sup>. Additional sort discriminators were 4',6'-diamidino-2-phenylindole (viable), low side scatter, and singlets (forward scatter pulse area vs width). Small DP thymocytes were identified as being forward scatter low as well as expressing high levels of CD4 and CD8. Only sorted populations that were >99% pure were used as template for microarray analysis. RNA was extracted using StrataPrep total RNA kits as recommended by the manufacturer (Stratagene, La Jolla, CA). Overall, purified cells from 16 different cell sorting days, each including 8–10 thymuses, were pooled to make a minimum of 5 μg of RNA template for further analysis.

Probe synthesis and chip hybridization

RNA was not amplified before use as a template to generate labeled probes. All procedures were conducted by the Molecular Genomics Core Facility at Memorial Sloan-Kettering Cancer Center (New York, NY). Biotin-labeled cRNA probes were prepared as recommended by the chip manufacturer (Affymetrix, Santa Clara, CA). Briefly, 5 μg of total cellular RNA were used to generate double-stranded cDNAs with oligo(dT) and SuperScript reverse transcription reagents (Invitrogen, Carlsbad, CA). The resulting dsDNA was used to prepare biotinylated cRNA probes with a BioArray High Yield RNA Transcription kit (Affymetrix). Biotin-labeled cRNA was fragmented and hybridized to the MG-U74A array for 16 h at 45°C as recommended by the manufacturer. Washing was performed using an automated fluids workstation, and the array was imme-
diately scanned on a Hewlett Packard GeneArray Scanner (Hewlett Packard, Palo Alto, CA).

Identification of genes expressed at each stage of development

The methods for analysis of the chip image were those recommended by the manufacturer (Affymetrix). Two primary parameters are of interest for each gene on the chip, namely, the absolute call (present, absent, or marginal) and the signal intensity. Calculation of these values was achieved as follows. The difference between probe pair (matched/mismatched) intensities was used to generate a discrimination score. If the discrimination score was higher than a default predefined threshold (τ 0.015), the gene was assigned a call of present; if it was lower, a call of absent or marginal was assigned. The one-sided Wilcoxon signed rank test was then used to determine a weighted mean of the discrimination scores for each probe set (i.e., for each gene). These results were then used to generate confidence limits that the combined score was different from the default threshold (τ). When the confidence interval indicated that the weighted mean was significantly different from τ (τ<sub>pval</sub> < 0.04), the gene product was designated as being present. A list of genes present in each stage was then prepared, and for these the signal intensity value was extracted. Signal intensity represents a quantitative assessment of the relative level of expression for each individual gene. This is determined by taking the log of the difference in intensity between each matched/mismatched probe pair (negative values are not used) and then calculating a weighted mean using the one-step Tukey biweight estimate. This mean value is then expressed relative to the mean of all genes found to be present on the chip, set to an arbitrary value (in this case, 500). Thus, genes that are present (τ<sub>pval</sub> < 0.015) and have a signal intensity of 1000 would be categorized as being expressed at roughly twice the mean level for all genes in that sample.

Identification of differentially expressed genes and transcription factors

The list of all genes present at one or more stage of differentiation (described above) was merged, and the signal intensities of adjacent stages of differentiation (e.g., DN1 vs DN2, DN2 vs DN3, etc.) were compared with identify genes that underwent changes of ≥2-fold at each developmental transition. This list was then annotated by submitting a batch query to the Affymetrix NetAffx website (https://www.affymetrix.com/analysis/query/interactive_query.jsp) using the probe set identification numbers corresponding to these differentially expressed genes. This annotated list was then filtered based on Gene Ontology (GO) Consortium designations related to transcription factor activity. Specifically, the following designations were used: GO biological process numbers 16481 (negative regulation of transcription), 45941 (positive regulation of transcription), 6350 (transcription), 6351 (transcription, DNA dependent), 6355 (regulation of transcription), 6366 (transcription from promotor), 6367 transcription factor from promotor; GO cellular component numbers 5634 (nucleus) and 5667 (transcription factor complex); GO molecular function numbers 3676 (transcription activator), 16565 (transcriptional repressor), 3676 (nucleic acid binding), 3677 (DNA binding), 3700 (transcription factor), 3712 (transcription coactivator), 3713 (transcription corepressor). Irrelevant genes emerging from these categories (e.g., RNA polymerases, histones, etc.) were manually removed from the list.

Real-time RT-PCR analysis

cDNA was synthesized from DNase-treated RNA from sorted thymocyte subsets (prepared as described above), using random hexamer primers and Superscript II reverse transcriptase (InVitrogen). Real time, quantitative PCR were performed in duplicate. A standard curve consisting of four 5-fold dilutions of thymus or brain cDNA was included in each experiment for each primer pair. PCRs were performed for 40 cycles on an ABI 7900 Sequence Analyzer (Applied Biosystems, Foster City, CA) using a SybrGreen master mix according to the manufacturer’s instructions. After the final cycle, automatic melting curve analysis was performed to ensure that only single PCR products were included in the data analysis. The final product was also analyzed by agarose gel electrophoresis to verify the size and integrity of the product and to ensure that it corresponded to the single peak seen in melting curve analysis (transcriptional repressor), performed using SDS2.0 software (Applied Biosystems). The minimum number of cycles required for detection above a threshold (cycle threshold, C<sub>t</sub>) was used to calculate the absolute cDNA amount in each sample by extrapolating C<sub>t</sub> values on the four-point standard curve. Expression of a housekeeping gene (glyoxaldehyde-guanine phosphoribosyltransferase; HPRT) was likewise calculated for each sample. The relative level of expression for each transcription factor gene was then calculated by dividing...
its cDNA value by the corresponding HPRT cDNA value. Primer sequences were as follows.

- **Bcl6**
  - Forward: CCGTACCCCTGTGAAATCTG
  - Reverse: AAGTCGCAGTTGGCTTTTGT

- **Ezh1**
  - Forward: AAAACGGAGCGGCATTATGCTA
  - Reverse: TGTGCACTGAGGGGGAAGTG

- **Ezh2**
  - Forward: TTCGTGCCCTTGTGTGATAG
  - Reverse: AGCATGGACACTGTGGTG

- **Egr1**
  - Forward: CAGGAGTGATGAACGCAAGA
  - Reverse: TGGGGATGGGTAAGAAGAGA

- **Klf7**
  - Forward: CAAGTGCTCATGGGAAGGA
  - Reverse: TGGTCAGACCTGGAGAAACA

- **Klf3**
  - Forward: AGAACCATCCTTCCGTATC
  - Reverse: GGTGCATTTGTACGGCTTTT

- **Irf7**
  - Forward: CCTCTTGCTTCAGGTTCTGC
  - Reverse: GCTGCATAGGTCCTCGTA

- **Irf5**
  - Forward: TTCCAGAAGGGCCAGACTAAT
  - Reverse: TGACATCAGGCCATTCTTCTC

- **BhlhB2**
  - Forward: AGCCGTGACTTGAAAGAGA
  - Reverse: TGGATGACTGGCACACAGTT

- **c-Fos**
  - Forward: ATCCTTGGAGCCAGTCAAGA
  - Reverse: TCCCAGTCTGCTGCATAGAA

- **c-Jun**
  - Forward: TAACAGTGGGTGCCAACTCA
  - Reverse: CGCAACCAGTCAAGTTCTCA

- **Rorc**
  - Forward: GCCCACCATATTCGACTCCTACATACCT
  - Reverse: TAAGTTGGCCGTCAGTGCTAT

- **Nab1**
  - Forward: GACCCACACAAAGAGGAGGA
  - Reverse: GGGCATTGTCCTTCACACAC

- **Nab2**
  - Forward: GAACCAGAGATGGTGCGAAT
  - Reverse: TTCCGGATCTCCTCTTCCTT

- **Il2ra**
  - Forward: AACGGCACCATCCTAAACTG
  - Reverse: CTGTGTTGGCTTCTGCATGT

- **Hprt**
  - Forward: ATCAGTCAACGGGGGACATA
  - Reverse: TTGCGCTCATCTTAGTGCTTT

### Results

#### Changes in transcription factor expression during inthrymphic lymphoid progenitor development

To reveal new candidates involved in transcriptional control of T lymphocyte developmental programs, we performed microarray analysis using Affymetrix high density mouse arrays (U74A). RNA was extracted from purified populations of inthrymphic lymphopoietic progenitors, including DN1, DN2, DN3, and preDP cells, as well as small, noncycling DP cells as lineage committed, postmitotic controls. The strategy for identification of transcription factors that might be important during specific developmental transitions is illustrated in Fig. 1. First, a list of all genes present (see Materials and Methods) in at least one progenitor stage was prepared. Next, this list was filtered to provide a list of genes that changed more than 2-fold on transition from one stage of differentiation to another. This list of all genes that change at one or more developmental transitions was then annotated by a batch query of the Affymetrix database. GO designations (www.geneontology.org) were then used to specifically identify genes from this list that are associated with transcriptional regulation. This final list of transcription factors that undergo >2-fold changes in expression levels at defined developmental transitions was used to generate the data shown in Figs. 2–5. In addition, genes that are expressed throughout all subsets but do not change >2-fold at any individual transition are presented in Supplemental Table I. Some of these may change >2-fold over multiple phases of the transition from DN1 to smDP. In the interest of space, we have not attempted to add the numerous additional figures that would be required to display such changes. For those interested, such findings can be extracted from Supplemental Table II, which provides the signal levels and present/absent calls for all genes described in all figures and tables of this article.

#### FIGURE 1

Strategy for bulk identification of transcription factors present during early thymocyte differentiation.

#### FIGURE 2

Transcription factors that are differentially expressed during the DN1 to DN2 transition. Values on the horizontal axis indicate fold changes in transcription factor expression between stages, as indicated. Only transcription factors with a >2-fold change are shown. Values for those factors that changed >10-fold are indicated next to the corresponding bar. Fold changes depicted as infinite (∞) indicate that the corresponding factor was only present in one of the two stages being compared; these genes are listed in Table I.

4 The on-line version of this article contains supplemental material.
functions, not all proteins generally considered to be transcription factors are necessarily revealed when filtered based on GO annotations. However, at the time of this writing, every effort has been made to ensure that all relevant genes are included, including by manual inspection of gene lists as described in Materials and Methods. It is also important that neither c-fos nor c-jun were identified by microarray analysis, either as genes that change during the differentiation process (Figs. 2–5) or as common genes (Supplemental Table I). This is not because these genes are not expressed in progenitor cells (Fig. 6), but rather they result from poor probe design and performance on the U74A chip; i.e., signal was detected in the matched oligos but was not significantly different in the single-base mismatch. Despite such occasional problems, the microarray analysis does provide, for the most part, reliable and informative results, as described in further detail below.

A few caveats regarding these data are warranted. First, all populations of progenitor thymocytes contain cells at different stages of cell cycle (19, 20), including smaller (G1-S phases) and more blastic (G2-M phases) cells, and the relative proportions of such cells may influence the gene expression outcomes for any individual phase. Likewise, because differentiation in the steady state thymus is not synchronous, each subset likely includes a proportion of cells that have already received signals to differentiate to the next stage, as well as cells that have not yet received such signals, and the relative proportions of these again may contribute to the gene expression patterns attributed to any single subset. Finally, without additional verification (such as by RT-PCR; see Validation of microarray results by analysis of known genes for selected genes in this category), it may be risky to generalize the fold changes and/or absolute expression levels described here. Rather, we believe that our results should be used as guidelines for more critical assessments of gene expression and/or validation of existing results.

Validation of microarray results by analysis of known genes

To most accurately assess gene expression patterns during intra-thymic T lymphocyte development, RNA extracted from purified cells was not amplified before probe synthesis. Purification of sufficient RNA template from rare DN1 and DN2 progenitors was therefore quite costly and very time consuming, making replicate microarray analysis an unattractive option. Therefore, to validate microarray results, two other types of confirmatory analyses were performed. In the first, microarray results for genes with well-documented changes in expression patterns during thymocyte differentiation were examined. A small panel of these is shown in Fig. 8. Consistent with what is already known (27), CD3e is not expressed at statistically significant levels in DN1 cells but increases in expression thereafter. Also consistent with previously published results (28), the invariant preTc component of the pre-TCR (16) is highest on DN3 cells and lowest on DN1 and DN2 cells. Likewise, the RAG-2 and c-kit expression patterns revealed by microarray analysis very closely parallel those previously reported by others (29, 30). CD25, which marks intermediate stages of DN differentiation (30–32), is found at significant levels only on DN2 and DN3 cells by microarray analysis, whereas CD4 is upregulated only at the DP stage. We have also reported other consistencies between these microarray results and other methods of analysis (33). In general, microarray results for genes known to be differentially regulated during T cell differentiation were remarkably consistent with what would be expected, suggesting that most transcription factors identified by this method (Figs. 2–5) are likely to be valid candidates for further investigation.

Validation of microarray results by real-time, quantitative RT-PCR

To specifically validate microarray data regarding potential transcriptional regulators of thymocyte differentiation, real time RT-PCR was performed for 12 selected transcription factors identified by microarray screening, as well as 1 additional control gene (CD25). The results of these analyses are shown in Fig. 6. Candidates for confirmation were selected to span the variety of patterns and expression levels found, i.e., genes that went down during differentiation, up during differentiation, or various combinations thereof, as well as genes that were relatively abundant vs those that were present but less abundant. In general, the
patterns of gene expression revealed by real time PCR were remarkably consistent with those found by microarray analysis. The only significant exception to this was Ezh2, which appears to remain high or increase slightly at the DN2, DN3, and preDP stages by microarray results, while decreasing over the same period by real time PCR. In other cases, such as Irf7, the patterns of expression were similar, but the overall fold changes were slightly different between real time and microarray results. Bhlhb2 and Nab2 exhibited similar trends by both microarray and real time PCR but differed slightly at a single transition (DN1/DN2 or DN2/DN3, respectively). Overall, however, both the patterns and relative fold changes appear to be very similar for both types of analysis. Together with the comparisons described earlier, these findings suggest that, in general, the microarray results reflect a fairly true assessment of gene expression and have the potential to reveal novel transcriptional candidates important in control of the intrathymic differentiation process.

Discussion

The differentiation of T lymphocytes in the postnatal thymus is initiated by signals originating from stromal elements that form the thymic microenvironment (reviewed in Ref. 22). The variety of signals assimilated by individual precursor cells are subsequently manifest as changes in transcriptional activity, many of which are, in turn, a consequence of changes in transcription factor expression. Thus, the identification of transcription factors that are expressed by all early thymocytes (Supplemental Table I), and especially those that fluctuate during the differentiation process (Figs. 2–5), represents an essential step in characterizing the biology of the lymphopoietic process. The advent of high density microarray analysis, as presented here, represents a high throughput method for accomplishing this goal. In particular, we have focused on the lymphopoietic stages of intrathymic differentiation, i.e., those stages that derive from non-self-renewing blood-borne progenitors and that generate the raw materials for positive and negative selection (i.e., smDP cells).

To accurately assess the expression of transcription factor genes in early thymocyte progenitors, RNA was isolated directly from cells and used without further amplification. Instead, numerous samples were pooled from multiple days of purification and cell sorting to make adequate RNA (5 μg) for the labeling and hybridization processes. Analysis of predicted patterns of expression (see...
The progenitors. The y-axis values represent relative fos or jun expression, compared with that of a housekeeping gene (HPRT), nominally defined as equal to 1.

Fig. 8) confirmed the usefulness of this approach, as the vast majority of genes for which expression patterns are known were corroborated by this analysis (additional data not shown). Further, real time quantitative PCR analysis for randomly selected transcription factors identified by the present studies also confirmed the validity of the results (Fig. 6).

In addition to the validations described above, comparison of transcription factors already known to be involved in the T cell differentiation process correlated extremely well with gene expression results by microarray. For instance, forced expression of Sfpi1 (PU.1) caused a block at the DN2/DN3 transition (34), indicating that down-regulation of this transcription factor is required for this transition to occur normally. Consistent with this, our data show that DN1 and DN2 cells express Sfpi1, but it is absent after DN3 (Table I). Deficiency in Gfi1 leads to developmental arrest at the DN2 stage (35), a stage at which our data indicate that it is substantially up-regulated (Fig. 2). In Tcf1 mutant mice (36), differentiation is impaired just before the DP stage, a stage at which our data indicate that Tcf1 is up-regulated (Table I). Consistencies were also noted between our data and published findings for Egrs (37) c-Myb (38), GATA-3 (39), Lmo1 (40), Heb (41), Runx1 (42), and many others. Overall, these consistencies together with the documentation provided by known genes (Fig. 8) and quantitative PCR (Fig. 6) indicate that the additional novel transcription factors revealed by this study are very likely to be authentic regulators of thymocyte differentiation. Although the sheer volume of factors revealed by these studies precludes a discussion of all of them, a few examples, together with their potential relevance, are discussed below.

Regulation of proliferation during early thymic lymphopoiesis

Bone marrow thymocyte progenitors expand 1 million-fold on entry into the thymus but ultimately withdraw from the cell cycle before maturation and export to the periphery (7). Expansion at later intrathymic stages (i.e., at the DN/DP transition) is associated with expression of components of the TCR (43). However, the regulation of proliferation at earlier stages is still largely uncharacterized. Of particular interest is the down-regulation of proliferation at earlier stages is still largely uncharacterized. Of particular interest is the down-regulation of proliferation status that occurs at the transition from DN2 to DN3 (19, 20). Evaluation of our microarray data revealed a profound decrease in Lyll associated with this transition (Table I). Breakpoint translocation fusng Lyll to the TCRβ locus is a cytogenetic abnormality
associated with ~20% of human T cell acute lymphocytic leukemias (44, 45), suggesting a role for Lyl1 in T cell proliferation. TCRβ gene rearrangements also initiate at the DN2/DN3 transition (11, 30), further strengthening the association and suggesting that mistakes in the somatic (VDJ) recombination process in developing T cells may contribute to aberrant proliferation and oncogenic transformation. Lyl1 is highly expressed early in T cell development (DN1 and DN2) but is completely repressed in DN3 cells (Table I), consistent with RT-PCR results of others (46). Thus, our data indicate the possibility that down-regulation of Lyl1 may represent a critical step in repressing cell cycle activity during VDJ recombination in developing lymphocytes.

### Transcriptional control of Notch signaling

The mammalian homologs of the *Drosophila* hairy-enhancer of split (Hes) genes encode at least seven basic helix-loop-helix transcription factors (Hes1–7). Hes gene products are major transcriptional mediators of Notch signaling (reviewed in Ref. 47). Notch plays an important role in T lymphopoiesis (reviewed in Ref. 48), and regulation of Hes genes can therefore impact T cell differentiation by regulating Notch signals. For instance, inactivation of either Notch1 or Hes1 results in similar arrest at early stages of intrathymic T cell differentiation (49, 50). In the present data, Hes1 was expressed in the DN1–3 stages, significantly down-regulated

### Table I. Transcription factor genes that turn on or off at specific developmental transitions

<table>
<thead>
<tr>
<th>DN1/DN2 Transition</th>
<th>DN2/DN3 Transition</th>
<th>DN3/precDP Transition</th>
<th>PreDP/smDP Transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turned OFF</td>
<td>Turned ON</td>
<td>Turned OFF</td>
<td>Turned ON</td>
</tr>
<tr>
<td>1190010I08Rik</td>
<td>483408P15Rik</td>
<td>2610110L04Rik</td>
<td>Creb1</td>
</tr>
<tr>
<td>4921515A04Rik</td>
<td>Atm</td>
<td>930025P16Rik</td>
<td>Egr1</td>
</tr>
<tr>
<td>4930565N07Rik</td>
<td>Bad1</td>
<td>930022110Rik</td>
<td>483408P15Rik</td>
</tr>
<tr>
<td>4930566A11Rik</td>
<td>Bns1I-pending</td>
<td>C630022N07Rik</td>
<td>Ldb3</td>
</tr>
<tr>
<td>Ankrd2</td>
<td>Ddx6</td>
<td>Creq</td>
<td>Kesh2</td>
</tr>
<tr>
<td>Adh1</td>
<td>Etx2</td>
<td>Chpb2</td>
<td>Lef1</td>
</tr>
<tr>
<td>Atp3</td>
<td>Fh12</td>
<td>Deaf1</td>
<td>Nab2</td>
</tr>
<tr>
<td>Baf1</td>
<td>Hivep2</td>
<td>Dnkk1</td>
<td>Nrl1</td>
</tr>
<tr>
<td>Bcl3</td>
<td>Hoxa7</td>
<td>Dl5x</td>
<td>Nrl2</td>
</tr>
<tr>
<td>C2ta</td>
<td>Ikdkg</td>
<td>Etk3</td>
<td>Nrl4</td>
</tr>
<tr>
<td>Cebp</td>
<td>Mud3</td>
<td>Gb1b</td>
<td>Pax6</td>
</tr>
<tr>
<td>Cd3dp</td>
<td>Mllt10</td>
<td>Hoxa6</td>
<td>Pbx3</td>
</tr>
<tr>
<td>Clock</td>
<td>Myh12</td>
<td>Hoxa7</td>
<td>Pouf6f1</td>
</tr>
<tr>
<td>Dri1</td>
<td>Myst2</td>
<td>Hoxb8</td>
<td>Rbpsuhl</td>
</tr>
<tr>
<td>Egr1</td>
<td>Nfatc2</td>
<td>Hoxd3</td>
<td>Sox15</td>
</tr>
<tr>
<td>Eif2ak3</td>
<td>Nfkbie</td>
<td>Irf3</td>
<td>Spib</td>
</tr>
<tr>
<td>Fos</td>
<td>Nr1h2</td>
<td>Lmo1</td>
<td>Tcef2</td>
</tr>
<tr>
<td>Gata6</td>
<td>Pkrox1</td>
<td>Lyl1</td>
<td>Trpc4</td>
</tr>
<tr>
<td>Hik2</td>
<td>Pla2g6</td>
<td>Mafg</td>
<td>Trpc4</td>
</tr>
<tr>
<td>Hlx</td>
<td>Rfx2</td>
<td>Mllt10</td>
<td>Zfp105</td>
</tr>
<tr>
<td>Hmgc2</td>
<td>Rxank</td>
<td>Nfatc2</td>
<td>Zfp46</td>
</tr>
<tr>
<td>Hoxa2</td>
<td>Sap18</td>
<td>Nhl3</td>
<td>Zfp60</td>
</tr>
<tr>
<td>Hoxa4</td>
<td>Sox1</td>
<td>Nhl1</td>
<td>Zfp60</td>
</tr>
<tr>
<td>Hoxa9</td>
<td>Sox7</td>
<td>Creg</td>
<td>Zfp42</td>
</tr>
<tr>
<td>Id3</td>
<td>Ta1</td>
<td>Rabb25</td>
<td>Zfp385</td>
</tr>
<tr>
<td>Irf4</td>
<td>Tcf20</td>
<td>Rfx2</td>
<td>Zfp46</td>
</tr>
<tr>
<td>Ir7</td>
<td>Tle4</td>
<td>Sfp1</td>
<td>Zfp46</td>
</tr>
<tr>
<td>Irf7</td>
<td>Zfp26</td>
<td>Sox3</td>
<td>Zfp46</td>
</tr>
<tr>
<td>Jnb</td>
<td>Zfp260</td>
<td>Sox7</td>
<td>Zfp46</td>
</tr>
<tr>
<td>Klf4</td>
<td>Zfp358</td>
<td>Ta1</td>
<td>Zfp26</td>
</tr>
<tr>
<td>Klf5</td>
<td>Zfp97</td>
<td>Tle4</td>
<td>Zfp26</td>
</tr>
<tr>
<td>Lef1</td>
<td>Zfp26</td>
<td></td>
<td>Zfp26</td>
</tr>
<tr>
<td>Lisch7-pending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lmo2</td>
<td>Zfp358</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lwe1</td>
<td>Zfp422</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nab2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nva1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nva1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pax5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pax6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pbx1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pbx3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prdm1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rbpsuhl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spib</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tce2b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tnfsf5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zfp105</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zfp36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zfp46</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* These genes are designated as infinite (−) fold changes in Figs. 2–5. A categorical summary of this table (i.e., by transcription factor family) can be found in Supplemental Table III.
upon the transition to preDP, and lost in smDP thymocytes. The expression of Hes1 thus strongly parallels published data regarding patterns of Notch1 expression (51) and signaling during the DN to DP transition (52). There is also a modest, <2-fold, up-regulation of Hes1 during the DN1 to DN2 transition (not shown), likewise in accordance with published data (53), again corroborating the validity of microarray results presented here. However, our data also provide new insights regarding the regulation of Notch signaling in T cell development. Hes1 acts as a transcriptional corepressor upon interaction with members of the Groucho family of proteins (reviewed in Ref. 54). One member of this family, Groucho-5 (also known as amino-terminal enhancer of split, Aes), is specifically up-regulated on transition to the DN3 stage and down-regulated again in preDP cells, mirroring the expression peaks for Hes1 (Table I). Of interest, Groucho-5 is again up-regulated at the smDP stage (Table I). Thus, down-regulation of Groucho-5 appears to correlate specifically with the DN to DP transition, where Notch activity is known to be required (55). The up-regulation of Groucho-5 at the smDP stage, where Hes1 is down-regulated (Table I) may seem counterintuitive. However, Groucho proteins also interact with a large variety of transcription factors other than Hes (54). Further, we find Hes6 (Supplemental Table I) and Notch3 (data not shown) are both expressed at the smDP stage, suggesting that Groucho-5 may alternatively interact with these. It is also possible that interaction of Groucho-5 with Hes6 (56) could interfere with the activity of Hes1, thus promoting terminal differentiation in DP cells similar to its role in the nervous system (57).

Cell cycle arrest and cell death or survival at the DP stage

Another cluster of transcriptional regulatory complexes that appears to be up-regulated at the transition to smDP (Fig. 5) includes the Krüppel-like factors (KLFs), the CBP/p300 transcriptional coactivator Cited2, and p300/CBP-associated factor (PCAF). A number of KLFs are up-regulated in smDP cells, including Klf3, 7, and 13 (Fig. 5). KLFs are heavily implicated in cell differentiation, cell cycle regulation, and cell survival/death decisions (reviewed in Ref. 58), all of which are characteristic of smDP cells (7). KLFs have already been shown to have various roles in cells of the T lineage, including differentiation, cell cycle withdrawal, and survival (for reviews, see Refs. 59 and 60). One particularly interesting factor in this family is KLF13. KLF13 has a role in the differentiation of effector T lymphocytes (61, 62). KLF13 transcriptional activity is regulated by the transcriptional coactivators CBP/p300 and PCAF (63), and PCAF and Cited2 are up-regulated simultaneously with KLF13 at the smDP stage (Fig. 5). Another factor, KLF7, has been implicated in exit from cell cycle by inducing expression of cell cycle inhibitors (64), many of which we find to be up-regulated in smDP cells (not shown). Thus, our data indicate that the activity of KLFs, and in particular KLF7 and/or 13 and their regulatory partners, may represent good candidates for further characterization of the poorly understood networks that control cell cycle arrest, homeostasis, and cell survival/death of thymocytes.

Overall, the data presented here provide numerous novel observations regarding the transcriptional control of thymic T lymphocyte differentiation. Our confidence in the data is such that a number of these are already being followed up in our laboratories, using transgenic and knockout approaches. However, the sheer volume of factors found is daunting, and thus we believe that these data provide a rich resource for gene discovery by others. In addition, these findings may be used to confirm gene expression analysis derived from other methods, such as RT-PCR and to break those findings down into discrete subset results when whole thymocytes have been used as template.


