Evolutionary Origins of Lymphocytes: Ensembles of T Cell and B Cell Transcriptional Regulators in a Cartilaginous Fish

Michele K. Anderson, Rashmi Pant, Ann L. Miracle, Xiao Sun, Carl A. Luer, Catherine J. Walsh, Janice C. Telfer, Gary W. Litman and Ellen V. Rothenberg

J Immunol 2004; 172:5851-5860; doi: 10.4049/jimmunol.172.10.5851
http://www.jimmunol.org/content/172/10/5851

References

This article cites 51 articles, 25 of which you can access for free at: http://www.jimmunol.org/content/172/10/5851.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
Evolutionary Origins of Lymphocytes: Ensembles of T Cell and B Cell Transcriptional Regulators in a Cartilaginous Fish

Michele K. Anderson, Rashmi Pant, Ann L. Miracle, Xiao Sun, Carl A. Luer, Catherine J. Walsh, Janice C. Telfer, Gary W. Litman, and Ellen V. Rothenberg

The evolutionary origins of lymphocytes can be traced by phylogenetic comparisons of key features. Homologs of rearranging TCR and Ig (B cell receptor) genes are present in jawed vertebrates, but have not been identified in other animal groups. In contrast, most of the transcription factors that are essential for the development of mammalian T and B lymphocytes belong to multigene families that are represented by members in the majority of the metazoans, providing a potential bridge to prevertebrate ancestral roles. This work investigates the structure and regulation of homologs of specific transcription factors known to regulate mammalian T and B cell development in a representative of the earliest diverging jawed vertebrates, the clearnose skate (Raja eglanteria). Skate orthologs of mammalian GATA-3, GATA-1, EBF-1, Pax-5, Pax-6, Runx2, and Runx3 have been characterized. GATA-3, Pax-5, Runx3, EBF-1, Spi-C, and most members of the Ikaros family are shown throughout ontogeny to be 1) coregulated with TCR or Ig expression, and 2) coexpressed with each other in combinations that for the most part correspond to known mouse T and B cell patterns, supporting conservation of function. These results indicate that multiple components of the gene regulatory networks that operate in mammalian T cell and B cell development were present in the common ancestor of the mammals and the cartilaginous fish. However, certain factors relevant to the B lineage differ in their tissue-specific expression patterns from their mouse counterparts, suggesting expanded or divergent B lineage characteristics or tissue specificity in these animals. The Journal of Immunology, 2004, 172: 5851–5860.

T and B lymphocytes represent an evolutionary innovation within the chordate lineage. To understand the genesis of this innovation, different components of the lymphocyte phenotype need to be considered. Of these properties, the best studied to date in nonmammalian organisms are involved in immune recognition and include the following: the rearranging Ag receptor genes, including the Igs and TCRs, the components of the rearrangement apparatus, such as recombinase-activating gene (Rag)-1 and Rag-2, and the polymorphic MHC genes that encode Ag presentation structures (MHC class I and II). However, lymphocyte development confers additional properties on these cells beyond their recognition specificities. Lymphocytes also acquire specific effector functions, proliferative mechanisms, and tightly regulated cellular life spans that link Ag recognition to immune responses, tolerance, and immunological memory. These properties in mammals are conferred and maintained by intricate transcriptional regulatory cascades with many developmental checkpoints. The extent to which this developmental programming is shared among the lymphoid-like cells of nonmammalian vertebrates is a separate criterion for evolutionary homologies and only recently has begun to be investigated (1–3).

The emergence of lymphocytes has been placed early in vertebrate evolution on the basis of recognition structures, cellular morphology, and tissue architecture. No lymphocyte-like cells have been described outside the vertebrates, even in invertebrate deuterostomes (e.g., echinoderms and urochordates). The vertebrates themselves are divided into two major groups: the more basal jawless vertebrates (agnathans) and the jawed vertebrates (gnathostomes). The agnathans, represented by lampreys and hagfish, have cells in several different tissues that morphologically resemble lymphocytes (4); however, extensive efforts to isolate rearranging Ag receptor genes from these animals have been unsuccessful (5, 6). Furthermore, the agnathans lack any recognizable spleen or thymus, which are major lymphoid organs in all gnathostomes. In sharp contrast, even the most basally divergent group of gnathostomes, the cartilaginous fish (epitomized by the sharks, skates, and rays, collectively termed elasmobranchs) are fully armed with the thymus, which are major lymphoid organs in all gnathostomes. In sharp contrast, even the most basally divergent group of gnathostomes, the cartilaginous fish (epitomized by the sharks, skates, and rays, collectively termed elasmobranchs) are fully armed with the capability for mammalian-like Ag recognition. They possess rearranging genes for Ig H and L chains, and α, β, γ, and δ TCR chains (7), and they also possess genes encoding Rag-1 (8), the mutagenic polymerase TdT (9), and polymorphic MHC proteins (10–12). The quantitative real-time RT-PCR, Spi, spleen focus-forming virus preferential integration (site); Ets, E26 (retrovirus) transformation-specific gene.

*Division of Biology, California Institute of Technology, Pasadena, CA 91125; †University of South Florida, Department of Pediatrics, Children's Research Institute, St. Petersburg, FL 33701; ‡Mote Marine Laboratory, Sarasota, FL 34236; and §Department of Veterinary and Animal Sciences, Paige Laboratory, University of Massachusetts, Amherst, MA 01003

Received for publication November 11, 2003. Accepted for publication February 27, 2004.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. This work was supported by National Aeronautics and Space Administration Grants NAG2-1370 and NAG2-1588 (to M.K.A. and E.V.R.), and National Institutes of Health Grant R37 AI2338 (to G.W.L.). M.K.A. was supported in part by the Institute for Biomedical Science, University of South Florida. C.A.L. received partial support through grants from the Vernal W. and Florence H. Bates Foundation and the Disney Wildlife Conservation Fund.

1 This work was supported by National Aeronautics and Space Administration Grants NAG2-1370 and NAG2-1588 (to M.K.A. and E.V.R.), and National Institutes of Health Grant R37 AI2338 (to G.W.L.). M.K.A. was supported in part by the Stowers Institute for Medical Research. A.L.M. was supported in part by the Institute for Biomedical Science, University of South Florida. C.A.L. received partial support through grants from the Vernal W. and Florence H. Bates Foundation and the Disney Wildlife Conservation Fund.

2 Current address: Sunnybrook and Women’s College Health Sciences Centre, 2075 Bayview Avenue, Room A340, Toronto, Ontario M4N 3M5, Canada.

3 Current address: National Exposure Research Laboratory, U.S. Environmental Protection Agency, 26 West Martin Luther King Drive, Cincinnati, OH 45268.

4 Current address: American Medical Group Association, 1422 Duke Street, Alexandria, VA 22314.

5 Address correspondence and reprint requests to Dr. Ellen V. Rothenberg, Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125. E-mail address: evroth@its.caltech.edu

6 Abbreviations used in this paper: Rag, recombinase-activating gene; GATA, zinc finger factor binding to DNA sequence GATA; EBF, early B cell factor; Pax, paired-box transcription factor; Runx, vertebrate homolog of Drosophila Runt; Q-PCR, quantitative real-time RT-PCR; Spi, spleen focus-forming virus preferential integration (site); Ets, E26 (retrovirus) transformation-specific gene.
homologies in at least these aspects of lymphocyte function make the cartilaginous fish extremely valuable as a launching point for any investigation to seek the traces of lymphoid evolutionary origins through analysis of more distantly related species.

Transcription factor genes can be used as powerful probes for evolutionary relationships over extended periods, because many of them belong to ancient families that are highly conserved even between the major deuterostome and protostome branches of the metazoa (multicellular animals). In principle, a conserved pattern of transcription factor usage could reveal relationships between the gnathostome lymphocyte program and some invertebrate lineages even if the recognition structures they use are different. In mammals, knockout studies and analysis of expression patterns have defined essential roles for many different transcription factors in lymphocyte development, and probable roles for several others (Fig. 1) (reviewed in Refs. 13–19). Zn finger factor binding to DNA sequence GATA (GATA)-3, paired-box transcription factor (Pax)-5, and early B cell factor (EBF)-1 have easily demonstrable lymphoid-specific roles within hemopoiesis. GATA-3 is essential for T cell development from the earliest stages. B cell development requires a transcriptional cascade involving the transcription factors E2A, EBF-1, and Pax-5. Whereas E2A also has roles in the development of other cell types, the requirements for EBF-1 and Pax-5 are highly restricted to B cells.

Other factors, which include vertebrate homolog of Drosophila Runt (Runx), PU.1, and Ikaros family members, also play critical roles in T or B cell development, although the effects of loss of function are rendered more complex, because these factors are also needed by stem cells and/or by more than one hemopoietic cell type. PU.1 is absolutely required for the development of B cells and macrophages, and is also needed for efficient production of T cell precursors (20, 21). Runx family members are essential for the generation of hemopoietic stem cells (22, 23), and play ongoing roles in lymphocyte and myeloid development as well. Multiple recent reports emphasize the importance of these factors in key T cell developmental checkpoints (24–30). Ikaros family members also are involved in lymphocyte development and/or function. Ikaros itself is needed for the development of all lymphocyte lineages (31) and for setting of T cell activation thresholds (32), whereas Aiolos is needed for proper B cell function (33). Helios is less well studied, but it is selectively expressed at high levels in stem cells and T cell precursors, suggesting a potentially restricted role in T cell development (34). PU.1, Runx, and Ikaros family members thus provide potentially ancestral bridges between lymphoid and nonlymphoid hemopoietic development.

The question we address in this study is the following: how much of the mammalian pattern of transcription factor usage described above is fundamental to the definition of lymphocytes, and conserved throughout all forms of lymphocyte development? This question is posed most strenuously by examining the most deeply divergent group from mammals that still possesses lymphocytes, namely the cartilaginous fish. The clearnose skate (Raja eglanteria) has been chosen as a model cartilaginous fish, because there is detailed information available about the tissue-specific and developmental regulation of multiple Ig isotypes and all four TCR genes (7, 9). Detailed knowledge of these genes provides an invaluable external standard to identify tissues that contain B cells and/or T cells, breaking the potential circularity that could otherwise affect arguments about evolutionary homology based on conserved combinations of transcription factors alone. This study examines the tissue-specific expression of skate homologs of lymphopoietic transcription factors and demonstrates that, although major components of the T cell and B cell gene regulatory networks observed in mammals are conserved in cartilaginous fish, other aspects of B cell developmental regulation differ markedly from that seen in mammals.

Materials and Methods

Experimental animals

Clearense skates were collected in near-shore waters off Anna Maria Island, FL and maintained in captivity at Mote Marine Laboratory (Sarasota, FL) under conditions previously described (35). Eggs laid following captive breeding of wild caught adults were incubated at 20°C to achieve reproducible development of embryos. Adult skates and embryos of known developmental age were sacrificed by immersion in seawater containing tricaine methane sulfonate (MS-222; Crescent Research Chemicals, Phoenix, AZ) at a concentration of 200 mg/L. Tissues were either processed immediately or frozen in liquid nitrogen and stored at −80°C until used.

Library screening

The GATA-3 probe was made by digesting the full coding length insert from a mouse GATA-3 cDNA plasmid. The EBF-1 probe, which contains the DNA binding domain, was generated by PCR using the primers (5′ to 3′) EBF-232, CGGAAATCCACCTCTTTCACTTTCC, and EBF-809, AGGGAGTAGCATGTTCCAGATAAG, and mouse spleen cDNA as template. The Runx probe was generated from the DNA binding domain of a mouse Runx1 cDNA. The gel-purified probes were radiolabeled and used to screen R. eglanteria spleen and/or thymus cDNA libraries, under low-stringency conditions as described previously (1). Clones were plaque-purified, and plasmids were generated by in vivo excision (Stratagene, La Jolla, CA).

Degenerate PCR

Short degenerate PCR primers were designed to amplify multiple family members for the Pax, GATA, and EBF based on conserved amino acid motifs in each DNA binding domain. EBF primers were as follows: 5′-EBF-DMNF: CAGGCGCGGCTGCAGAACATGG/TG, and 3′-EBF-DMNF: CAGGCGCGGCTGCAGAACATGG/TG. GATA primers were as follows: 5′-degGATA-YHKMN and 3′-degGATA3 (36). Pax primers were as follows: 5′-Pax-PCDI and 3′-Pax-WIEIR (37). Templates were 1/100 dilutions of first-strand cDNA generated from 10 µg of RNA from skate brain or skate spleen. The PCR products were amplified, cores were taken of the appropriately sized bands, and secondary degenerate PCR was performed as previously described (37). The PCR products were gel-purified and eluted using the Qiaex II kit (Qiagen, Valencia, CA), phenol/chloroform extracted, ethanol precipitated, and cloned into the pGEM-T Easy vector (Promega, Madison, WI).

**FIGURE 1.** Transcription factor expression and roles in mammalian lymphocyte development. This figure summarizes the lineage and developmental stage during which each factor is expressed; the squares show the first developmental stage that results in serious impairment when the factor is absent. Absence of a square in any of the categories indicates that knockout phenotype is unknown, with the exception of GATA-1, which is required for the erythroid lineage but not the lymphoid or myeloid lineages. Question marks indicate a lack of expression data in the published literature. See introduction and Discussion for references.

<table>
<thead>
<tr>
<th>Transcription Factor</th>
<th>T Lineage Early/Late</th>
<th>B Lineage Early/Late</th>
<th>Myeloid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ikaros</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Aiolos</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Helios</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>PU.1</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Spi-B</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Spi-C</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GATA-1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GATA-3</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>EBF-1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pax-5</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Runx-2</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Runx-3</td>
<td>+/+</td>
<td>+/+</td>
<td>-</td>
</tr>
</tbody>
</table>

Downloaded from http://www.jimmunol.org/ by guest on April 14, 2017
Sequence analysis

DNA sequencing was performed using the ABI Prism Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA). DNASTAR (Hichi, San Francisco, CA) and ClustalW software (38) were used to analyze the sequences and generate the amino acid alignments, respectively.

GenBank accession numbers

GenBank accession numbers are as follows: ReGATA3, AY079198; ReEFB1, AY079197; ReGATA1, AY395710; RePax5, AY395711; ReRunx2, AY444494; ReRunx3, AY444493; and RePax6, AY577906.

Real-time PCR expression analysis

First-strand cDNA was generated from total RNA from each tissue from mice (strain C57BL/6) or skates using the RNAzol B (Leedco Medical, Houston, TX) and Superscript RT II (Invitrogen, San Diego, CA) methods as described previously (1, 9). Quantitative real-time RT-PCR (Q-PCR) were run using first-strand cDNA, specific primers, and SYBR Green kit reagents (Applied Biosystems, Foster City, CA). The primers used are listed in Table I, except for the skate PU.1, spleen focus-forming virus

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>ReEFB-1</td>
<td>CAC AAG CAG CTT TCC TCC</td>
</tr>
<tr>
<td>ReGATA-3</td>
<td>TCG GGA GAG CTT TCC</td>
</tr>
<tr>
<td>ReEfkaros</td>
<td>AGC ACT CCT ACT ACT</td>
</tr>
<tr>
<td>ReElarios</td>
<td>GCA AAG TCG TGG AAG CTT</td>
</tr>
<tr>
<td>ReHelios</td>
<td>TGG CAT ACT TAG GAT GCG CTT</td>
</tr>
<tr>
<td>ReAiolos</td>
<td>TGC AGCCAA AGA GAA TCC AAA</td>
</tr>
<tr>
<td>ReGATA-1</td>
<td>TGA CAT GCT CAC CCT TGT</td>
</tr>
<tr>
<td>ReRunx-2</td>
<td>CCA AGA CCT AGA AGA CAT AGA CA</td>
</tr>
<tr>
<td>RePax-5</td>
<td>GGT GGA TAA AAT CGC CTA C</td>
</tr>
<tr>
<td>ReGATA-3</td>
<td>GCG AAT CTC CCA GGC AAA C</td>
</tr>
<tr>
<td>RePax-6</td>
<td>GCG GTG ATG AGC GGT GCA</td>
</tr>
<tr>
<td>ReRunx-3</td>
<td>TTC GAT GTG GCA TCC TCT</td>
</tr>
<tr>
<td>MmGATA-1</td>
<td>AGC TAT AGA GGG AAT GAT GGA ATG</td>
</tr>
<tr>
<td>MmGATA-3</td>
<td>TGT CCA AGC GGC GCA TCA</td>
</tr>
<tr>
<td>MmEFB-1</td>
<td>GAG GTG GTG TGT GCA TCC CAA</td>
</tr>
<tr>
<td>MmPax-5</td>
<td>GTA CAG CAC AGG ATT TTT CCA</td>
</tr>
<tr>
<td>MmSpi-C</td>
<td>GCT CCT CCT CAC GTC TCC</td>
</tr>
<tr>
<td>MmSpi-B</td>
<td>CAG GTG GGC CAG GCT</td>
</tr>
<tr>
<td>MmIkaros</td>
<td>GAA GGC AGC AGT AAC</td>
</tr>
<tr>
<td>MmEfkaros</td>
<td>TGG GGA GAG ACA GAT</td>
</tr>
<tr>
<td>MmElarios</td>
<td>ATG CAT CTG TCG</td>
</tr>
<tr>
<td>MmHelios</td>
<td>GTT CAT GCG CAG TTG</td>
</tr>
<tr>
<td>MmRunx-1</td>
<td>TGC CCA ACA GGA CCC</td>
</tr>
</tbody>
</table>

Re, Skate; Mm, mouse.

Results

Identification of skate transcription factor family members

The goal of this study was to test for equivalence between the skate and mammalian systems at three distinct levels: 1) the existence of transcription factor family members with orthologous structures (coding sequence conservation for individual genes); 2) the expression in vivo of a given transcription factor in similar patterns relative to sites of Ig or TCR expression (regulatory sequence conservation for individual genes); and 3) the existence of Ig- or TCR-expressing sites that coexpress corresponding groups of transcription factors (conservation of regulatory programming for cell types). None of these types of conservation could be assumed a priori. In particular, it could not be assumed that the member of a transcription factor family used in a skate lymphocyte population would necessarily be the one that was most structurally similar to the member of the family that is used in mammalian lymphocytes. Therefore, we sought to isolate from skate tissues not only plausible orthologs of mammalian hemopoietic transcription factors, but also any other family members that might be expressed in similar tissues.

Heterologous cross-hybridization and degenerate PCR strategies were used to recover members of the GATA, EBF, Runx, and Pax transcription factor families from the clearnose skate, R. eglanteria. Full-length cDNAs or partial-length PCR products were sequenced, and the predicted amino acid sequences were used to create alignments suitable for phylogenetic analysis. Each skate amino acid sequence was aligned with sequences from other vertebrate family members, using invertebrate sequences as outgroups. For the full-length sequences, these alignments were used to generate neighbor-joining trees from which orthology was assigned. Orthology is defined as homology between two genes separated only by speciation, whereas paralogy refers to genes separated by gene duplication. For example, mouse GATA-3 and human GATA-3 are orthologs, whereas mouse GATA-3 and mouse or human GATA-1 are paralogs. The phylogenetic analyses will be presented in full in a separate report. For the partial-length sequences, diagnostic amino acids for vertebrate orthologs were compared with the skate sequence to assign orthology. The full-length sequences are presented in Clustal alignments with their murine orthologs in Fig. 2, A–C. A diagram summarizing the percent identities by functional domain between skate and mouse orthologs of PU.1, GATA-3, EBF-1, and Runx3 is shown in Fig. 2D. This figure shows the strikingly high levels of amino acid identity that exist between the DNA binding domains of orthologous mouse and skate transcription factors, whereas areas outside the DNA binding domains vary in their conservation. Although we have previously described the domain conservation of PU.1, it is included in this figure as well for comparison.

Recovery of skate GATA family members

To obtain a potential skate GATA-3 homolog, we screened a skate spleen cDNA library with a probe encompassing the entire mouse GATA-3 coding region, and recovered one clone with an open reading frame of 444 predicted amino acids. This predicted protein shares extensive sequence identity with mammalian GATA-3 (Fig. 2B), as summarized on a domain-by-domain basis in D, including...
FIGURE 2. A–C, Clustal alignment of skate EBF-1 (A), GATA-3 (B), and Runx3 (C), with their murine orthologs. D, Domain-based analysis of four of the full cDNAs recovered from the skate as compared with their mouse orthologs. Domains with functions known in the mouse ortholog are marked with color, and white boxes indicate unknown function. Black, DNA binding domain; red, protein-protein interaction domain; pink, transactivation domain; blue, alternative start site. Numbers indicate percentage of amino acid identity between mouse and skate orthologs, with each gap counted as a mismatch. WA, Weak activation; SA, strong activation; PEST, protein-protein interaction domain; TA, transactivation domain; ZF, zinc finger; NLS, nuclear localization signal sequence; HLH, helix-loop-helix; prox, proximal promoter-driven 5'-end; NMTS, nuclear matrix targeting sequence; Gro/TLE, groucho binding motif; DBD, DNA binding domain. The scale for each factor is different to highlight different amounts of detail, and is indicated beneath each box diagram. Accession numbers for published sequences used in the alignments are as follows: MmEBF1, NM_007897; MmGATA3, NM_008091; MmRunx3, NM_019732; RePU.1, AF320627; and MmPU.1, NM_011355.
100% amino acid identity in the DNA binding domain. A PCR product representing a fragment of skate GATA-1 was recovered by degenerate GATA family primers using brain cDNA as a template. Although the sequence is short, several residues are diagnostic of GATA-1 orthology, as presented elsewhere. Multiple rounds of degenerate PCR using spleen and brain cDNA as template produced additional GATA-3 and GATA-1 products, but no additional members of the GATA family were recovered from these adult tissue cDNAs.

**EBF and Pax family members in the skate**

A mouse EBF-1 probe was used to screen the skate spleen cDNA library, and yielded a single clone with an open reading frame of 591 predicted amino acids. This sequence is over 95% identical with the EBF-1 amino acid sequences of mouse EBF-1, with 100% amino acid identity in the DNA binding domain, the HLH dimerization domain, and the region between them (Fig. 2, amino acid identity in the DNA binding domain, the HLH dimerization domain, and the region between them (Fig. 2, A and D). Degenerate PCR also was performed using EBF family-specific primers and skate spleen and brain cDNA. Three different products with significant identity to the EBF family were recovered using brain cDNA as template. The first was identical with a portion of the EBF-1 gene described above, and the other two were designated EBF-B and EBF-C. The orthology of EBF-B and EBF-C are thus far unclear, but the skate EBF-1 gene analyzed in this study appears by sequence analysis to be a true ortholog of vertebrate EBF-1 genes. Details of the characteristics of the three skate EBF family members will be presented in a separate report.

To obtain skate Pax homologs, we took advantage of the phylogenetically deep association between Pax gene expression (including members of the Pax2/5/8 family) and CNS development. Two Pax family member PCR products were recovered by degenerate PCR. One PCR product, recovered from spleen cDNA, appears to represent an ortholog of vertebrate Pax-5, whereas the other, obtained from brain cDNA, is an ortholog of Pax-6. Pax-6 is a member of a different Pax gene subfamily, and is not known to be expressed in mammalian lymphocytes, but instead plays a role in eye development throughout the bilaterians. Skate Pax-6 was used in this study primarily as a control for nonhemopoietic expression. Because only partial coding sequences of these genes were recovered as short PCR products, they are not included in the domain analysis of Fig. 2D.

**Skate Runx family members**

Screening of a skate thymus library resulted in the isolation of two different Runx family cDNAs. One of these represents a full-length ortholog of Runx3, whereas the other is a partial cDNA most related to vertebrate Runx2 genes. Fig. 2, C and D, shows that there is 100% amino acid identity between the proximal 5' end of mouse and skate Runx-3 and 94% amino acid identity in the Runx DNA binding domain, whereas the other domains are less conserved with the exception of the canonical Groucho/TLE binding motif at the 3' end. Interestingly, exon 6 (as defined by the Runx-3 mouse structure) is missing from the skate cDNA. In the mouse, exon 6 associates with Sin-3 and is involved in CD4 silencing. It is unclear thus far whether the skate cDNA represents a splicing variant or indicates that this exon has been lost from the gene locus. Repeated screens of thymus and spleen cDNA libraries have thus far failed to identify a skate ortholog of Runx1. This is surprising, because in a cDNA library made from murine thymocytes, >10 times more Runx1 clones were recovered than GATA-3 clones (M. K. Anderson, C. C. Tydell, E.-S. D. David, and E. V. Rothenberg, unpublished data); however, the possibility that a skate Runx1 also exists cannot be ruled out at this time.

**Conservation of lymphoid expression patterns in skate orthologs of murine lymphocyte transcription factors: T cell expression of GATA-3 and B cell expression of Pax-5**

The results of these cloning studies indicate that the skate possesses orthologs of many of the key transcription factors that control lymphocyte development in mammals. To determine whether these factors are expressed in sites containing lymphocytes, the expression levels of these factors were quantified in embryonic and adult skate tissues that have been characterized previously for their expression of TCR and/or Ig molecules diagnostic of the presence of T and B lymphocytes, respectively (9), as summarized in Table II. The major sites of lymphoid receptor expression in the adult skate are the thymus (T $\gg$ B), spleen (T and B), Leydig (B $\gg$ T) and epigonal organ (B $\gg$ T). The Leydig and epigonal organs are also rich in myeloid-type cells (39). These studies indicate that the Leydig is richer in lymphocytes (B $>$ M), whereas the epigonal organ is richer in granulocytes (M $>$ B). In addition to these major sites, some low-level TCR expression is also found in intestine, rectal gland, and liver.

Fig. 3 presents a side-by-side analysis of two GATA family members and two Pax family members in adult lymphoid (Ig and/or TCR$^+$) and nonhemopoietic tissues of the mouse and skate, respectively. GATA-3 in the mouse is expressed in several nonhemopoietic tissues in addition to T cells, under control of distinct sets of cis-regulatory sequences (40, 41). Thus, GATA-3 expression in the skate would not be expected to be confined solely to TCR$^+$ cells. However, if its strong association with T cell development is conserved, then it should be expressed in the strongly TCR$^+$ sites of thymus and spleen, in preference to other hemopoietic sites. Q-PCR analysis shows that GATA-3 is indeed expressed at high levels in the skate thymus and spleen with a notable lack of expression in the Leydig and epigonal organs (Fig. 3). Thus, as in mammals, this gene has a hemopoietic role in skate that is correlated with the presence of T cells as opposed to other hemopoietic cell types. This is a family member-specific role, because the skate sequence most closely related to GATA-3, namely GATA-1, is expressed in a different pattern.

A correspondingly cell type-specific role is played in the mouse hemopoietic system by the Pax-5 gene (Fig. 3). There, its role is tightly associated with B cell development. If Pax-5 is used this way in the skate, it should be expressed in a different hemopoietic pattern from skate GATA-3, with a bias toward sites of Ig expression. Q-PCR analysis shows that skate Pax-5 is expressed by far

**Table II. Skate tissue sites of Ag receptor expression**

<table>
<thead>
<tr>
<th>Tissue</th>
<th>IgHC</th>
<th>TCR</th>
<th>Major site</th>
<th>Minor site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embryo</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spleen</td>
<td>IgM, IgX, IgX</td>
<td>α, β, γ, δ</td>
<td>B cells</td>
<td>B cells</td>
</tr>
<tr>
<td>Thymus</td>
<td>IgX</td>
<td>α, β, γ, δ</td>
<td>B cells</td>
<td>B cells</td>
</tr>
<tr>
<td>Intestine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>IgX</td>
<td>α, β, γ, δ</td>
<td>T cells</td>
<td>T cells</td>
</tr>
<tr>
<td>Leydig</td>
<td>IgX</td>
<td>B cells</td>
<td>B cells</td>
<td>B cells</td>
</tr>
<tr>
<td>Gonad</td>
<td>IgX</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spleen</td>
<td>IgM, IgX, IgX</td>
<td>α, β, γ, δ</td>
<td>T, B cells</td>
<td>T, B cells</td>
</tr>
<tr>
<td>Thymus</td>
<td>IgM, IgX, IgX</td>
<td>α, β, γ, δ</td>
<td>T cells</td>
<td>T cells</td>
</tr>
<tr>
<td>Intestine</td>
<td>IgM, IgX, IgX</td>
<td>β, δ</td>
<td>T cells</td>
<td>T cells</td>
</tr>
<tr>
<td>Muscle</td>
<td>IgX</td>
<td>β, δ</td>
<td>T cells</td>
<td>T cells</td>
</tr>
<tr>
<td>Rectal</td>
<td>IgX</td>
<td>β, δ</td>
<td>T cells</td>
<td>T cells</td>
</tr>
<tr>
<td>Liver</td>
<td>IgX</td>
<td>β, δ</td>
<td>T cells</td>
<td>T cells</td>
</tr>
<tr>
<td>Leydig</td>
<td>IgX</td>
<td>B cells</td>
<td>B cells</td>
<td>B cells</td>
</tr>
<tr>
<td>Epigonal</td>
<td>IgX</td>
<td>B cells</td>
<td>B cells</td>
<td>B cells</td>
</tr>
</tbody>
</table>
FIGURE 3. Comparison of mouse and skate expression patterns of GATA and Pax family members. First-strand cDNA was synthesized from total RNA of each tissue and used as template for Q-PCR using the SYBR Green to detect the products. The difference between a normalization product (GAPDH for the mouse and 28S rRNA for the skate) and the experimental product used to calculate the initial transcription factor transcript relative to the normalization transcript assuming a 1.9-fold amplification efficiency per cycle. The resulting values were divided by the highest value in each series to give percentages of maximum, which indicate the relative levels of expression in each tissue as compared with the other tissues examined. The order of the tissues are arranged to maximize homology; however, the lymph node and bone marrow are found only in the mouse and not in the skate, whereas the Leydig and epigonal organs are found only in the skate and not in the mouse. Expression levels in this figure are presented on a linear y-axis scale.

the most strongly in the spleen in the adult, with much lower expression in the epigonal, Leydig, and nonhemopoietic sites such as brain. Skate Pax-6, conversely, is not expressed at any appreciable level in the thymus or spleen, but instead is found primarily in the brain. Skate Pax-6, conversely, is not expressed at any appreciable level in the epigonal, Leydig, and nonhemopoietic sites such as rectal gland, or thymus (data not shown). However, these family members identified in our screen, EBF-B and EBF-C, are expressed in all skate tissues, and in thymus, in addition to being expressed at higher levels in thymus and bone marrow. By analogy to mammals (44), adipose tissue may be a source of EBF expression within these other tissues (note the lower scale for EBF expression in Fig. 4B). Two other skate EBF/Olf family members identified in our screen, EBF-B and EBF-C, are also expressed in Leydig and epigonal, and in kidney, rectal gland, or thymus (data not shown). However, these family members appear to be less closely associated with B cell development than EBF-1, because they are expressed most strongly in the heart and very poorly in the major Ig expression site, the spleen.

Patterns of transcription factor expression in skate hemopoietic tissues: T cell- and B cell-associated clusters

Extension of the expression analyses to the PU.1/Ikaros transcription factor family (1) (see next section). The skate PU.1 and Spi-D genes are expressed preferentially in the adult Leydig and epigonal organs, with lower expression in spleen and virtually none in thymus. Thus, individual hemopoietic transcription factors are regulated in a tissue-specific manner in the skate hemopoietic system. Furthermore, in the cases of GATA-3 and Pax-5, the overlapping patterns of skate transcription factor expression with TCR or Ig expression are consistent with conserved developmental associations with the T and B lineages, respectively.

Runx3 plays an important role in a specific stage of mouse T cell development (24, 25).

Ikaros family members, in contrast, are found to be expressed strongly in all skate tissues with TCR and/or Ig expression, namely spleen, thymus, Leydig, and epigonal (Fig. 4A). This is of interest, because in mammals, B cells depend completely on Ikaros for development and on Aiolos for mature function, and T cells depend on both of these factors and possibly Helios as well for partially overlapping roles (33, 34, 42, 43). Thus, as in the mouse, the Ikaros family in the skate shows a general pattern of expression consistent with panlymphoid or panhemopoietic development.

The relationships of transcription factors with skate Ig+ cells reveal some complexity. As shown in Fig. 4B, one PU.1/Spi family member, namely Spi-C, is highly restricted in its expression to the spleen. This is similar to Pax-5, which is also expressed at its highest levels in the spleen. These two transcription factors from different families thus may share an association with B cells. However, they do not necessarily have roles in all B cell subtypes at all stages of B cell development. As discussed below, the levels of Pax-5 and Spi-C expression do not scale simply with levels of Ig expression; they are expressed at much lower levels in the IgM-rich Leydig organ than in the spleen. In mammals, Pax-5 expression is shut off in terminally mature B lineage plasma cells. Thus, even a strict conservation of the role of Pax-5 in B cell development would not require all Ig+ cells to express Pax-5 at all times. Another candidate B cell factor, EBF-1/Olf-1, is expressed in all the Ig+ tissues. However, it also is expressed very broadly among nonhemopoietic tissues and in TCR+ tissues such as rectal gland and thymus. By analogy to mammals (44), adipose tissue may be a source of EBF expression within these other tissues (note the lower scale for EBF expression in Fig. 4B). Two other skate EBF/Olf family members identified in our screen, EBF-B and EBF-C, are also expressed in Leydig and epigonal, and in kidney, rectal gland, or thymus (data not shown). However, these family members appear to be less closely associated with B cell development than EBF-1, because they are expressed most strongly in the heart and very poorly in the major Ig expression site, the spleen.
Developmental use of hemopoietic transcription factors in the skate

To determine whether the skate hemopoietic transcription factors are used during the differentiation of TCR$^+$ and Ig$^+$ cells, the lack of surface antigenic markers for various differentiation states required us to study tissues in which development is occurring on a known temporal basis. The first wave of lymphocyte development was therefore examined in the skate embryo. Skate embryos develop in egg cases for 12 wk before hatching (35), and hemopoiesis occurs in a burst, starting at 8 wk of embryonic development (9). At 8 wk, TCR expression is primarily restricted to the thymus, with only low levels detected in the spleen, whereas Ig expression occurs primarily in the spleen (see Table II). The Leydig organ, epigonal (gonad) organ, liver, and thymus also exhibit low levels of expression, primarily of the distinct IgX isotype (45, 46), which is expressed at high levels in germline-joined form only in the embryo (9).

In 8-wk-old embryos, GATA-3 is localized primarily to the thymus (Fig. 5) and correlates perfectly with TCR expression, which is found at significant levels only in the embryonic thymus. In contrast to the adult, GATA-3 is expressed in the embryonic thymus at nearly 10 times the level found in embryonic spleen.
Similarly, the prevalence of Helios and Aiolos transcripts is much higher in the thymus than in other hemopoietic tissues, as contrasted with the adult pattern, where they are more evenly distributed between thymus and spleen (Fig. 4A). Runx3 has an even more thymus-restricted pattern of expression in the embryo than in the adult. Thus, the first wave of lymphocyte development involves strong intrathyrmic participation by Ikaros family members, and specific involvement of the Runx3 (but not Runx2) factor. These results also suggest that much of the extrathyrmic hemopoietic expression of Helios in adults correlates with the emigration of T cells from thymus to periphery. The embryonic data thus suggest that, in addition to GATA-3, Helios and Runx3 also may have specific roles in T cell development.

Embryonic expression patterns of the PU.1/Spi family members confirm primary sites of expression for these factors in the Leydig and epigonal organs, from the first wave of hemopoiesis onward. However, one difference between embryonic and adult expression patterns of these factors relates to the presence of low but significant levels of both PU.1 and Spi-D in the embryonic thymus (Fig. 5). The absence of these factors from the adult thymus could be due to a shift in the developmental role of the organ, a shift in the T cell developmental program, or the expected dilution of the earliest types of T cell precursors by accumulated mature TCR+ cells in the adult organ. The last possibility is of interest, because in the mouse, PU.1 is expressed normally by the earliest T cell precursors and is functionally critical for fetal T cell development in particular (20, 21, 37, 47).

Both Pax-5 and GATA-1 exhibit differences in expression between the embryo and the adult. Pax-5 expression is higher in embryonic Leydig and epigonal than in embryonic spleen, whereas it primarily occurs in the spleen in the adult. Conversely, GATA-1 is higher in spleen than in the Leydig and epigonal organs in the embryo, but is more evenly distributed among these three tissues in the adult. In contrast, Spi-C is expressed primarily in the spleen in both the embryo and the adult. It should be noted that there are fewer nonhemopoietic tissues represented in the embryo expression profiles than in the adult profiles. Therefore, expression in a higher percentage of the tissues in the embryo than in the adult does not necessarily mean that the factor is expressed more broadly in the embryo.

Discussion

This report presents compelling evidence that most of the transcription factors involved in lymphocyte development in mice and men were available to play similar roles in lymphocyte development in the last common ancestor of the jawed vertebrates. We show in this study that the GATA, EBF, Ikaros, Pax, Runx, and PU.1 family members that are used selectively in mammalian lymphocyte development have orthologs that are present in a modern representative of the most ancient class of jawed vertebrates. In most cases, their expression patterns indicate conserved regulation. In general, the expression patterns of individual skate transcription factors exhibit linkages to the presence of TCR + or Ig + cells consistent with conserved cell type specificity between the skate and the mouse. Furthermore, coexpression patterns of certain transcription factor sets, and the shifts in these coexpression patterns in the embryo vs the adult, together indicate conserved roles in lymphocyte development as well as in mature lymphocyte function. Many different levels of regulation are involved in hemopoietic fate choice, including posttranslational and posttranscriptional modifications that are not monitored in this study. However, the tissue-restricted expression of the majority of the factors examined in this study indicate tissue-specific roles. These factors, if not active at all times in all tissues, are at least transcribed in response to regulatory inputs and poised for activity in response to specific signals, in a combinatorial and tissue-specific manner that is very similar to that seen in mammals. Our results thus suggest that the transcription factors described in this study constitute a basic framework for the programs directing T and B cell development, one that was likely established by the time that the rearranging Ag receptors appeared over 450 million years ago.

One of the aims of finding multiple members of each transcription factor family described in this report was to ascertain whether recruitment of specific family members to specific lymphocyte lineages was ancestral to the jawed vertebrate lineage. Different GATA family members, for example, might have been independently recruited to the T cell lineage in early vertebrates. Our results indicate that 1) there is clear orthology between jawed vertebrate transcription factors, indicating that the divergence of these factors occurred before the separation of the cartilaginous fish from the lineage leading to the mammals, and 2) that the same family members are used in each lineage among the factors tested in this study. Although this does not prove ancestral recruitment, it is the most parsimonious explanation. Furthermore, conservation of function is supported by the extremely high amino acid identity within the DNA binding domains, indicating the ability to bind the same sequence. These issues will be more fully explored in a separate report.

The expression of key transcription factor combinations correlates with Ig and/or TCR over a wide range of tissues, including some organs that lack homology between cartilaginous fish and mammals, and also persists across shifting expression sites during development. In particular, the consistent expression of GATA-3, Runx3, Ikaros, Aiolos, and Helios in the skate thymus of both embryos and adults provides strong evidence that these factors were involved in an ancestral regulatory program for thymic T cell development. Both TCR and Ig expression shift toward the spleen from embryo to adult, indicating a migration of mature lymphocytes to the spleen from their sites of development. Our studies show that the transcription factors associated with TCR expression also shift from thymus to spleen in a comparable way, indicating that these factors are used in both developing and mature T cells.

B cell development is more complex for several reasons. First, although Ig + cells, like TCR + cells, are found in the spleen, the other primary sites of Ig transcript expression (the Leydig and epigonal organs) are unique to the elasmobranchs. Furthermore, the embryonic spleen contains IgM+ cells, whereas cells in the embryonic Leydig and epigonal organs express a divergent Ig H chain (IgX) which is transcribed from germline-joined loci (9). However, in the adult, the spleen, the epigonal organ, and Leydig organ all contain both IgM-expressing and IgX-expressing cells. Analyses of Rag-1 and TdT expression have not detected enrichment in any particular site of Ig expression in the skate (9), although in the nurse shark Rag-1 and TdT have been found to be preferentially expressed in epigonal organ and not in the spleen (48). Therefore, mature B cells may exist in the Leydig and epigonal organs, which are also candidates for ongoing sites of B cell development in the skate. In the embryo, Pax-5 expression maintains a strong association with Ig-positive cells, consistent with a conserved role. However, the Ig-expressing cells in the adult epigonal and Leydig organs do not express Pax5. The shift of Pax-5 expression predominantly from Leydig and epigonal in the embryo to spleen in the adult could thus reflect lineage-specific or stage-specific roles in the different classes of B cells. Several possibilities include the following: 1) selective expression of Pax-5 by specific sublineages of B cells, such as IgX cells, or IgM lineage precursors before migration to the spleen, or 2) shutoff of Pax-5 in all terminally differentiated plasma cells (as in mammals), with the
implication that Igα Pax-5− plasma cells comprise the predominant B cell population in the Leydig and epigonal of the adult. The latter possibility would be in accord with the very high levels of expression of Ig transcripts reported previously in these sites (9). More detailed studies at a single-cell resolution will be required to sort out these possibilities.

In addition to Pax-5 in the spleen, Leydig, and epigonal organs, cells in each of these sites also have access to other factors orthologous to genes used in mammalian B cell development, such as EBF, Ikaros family members, and PU.1/Spi family members. This is a prerequisite for the operation of similar gene networks to direct B cell development in skate and mammal. However, some differences are evident. EBF-1, although present in all Ig-expressing sites, is not particularly prominent in any of them. Also, whereas Pax-5 expression appears to correlate with PU.1 expression in the embryonic skate, this is not the case in the adult. As Pax-5 expression shifts to the spleen, PU.1 and Spi-D remain most enriched in Leydig and epigonal. In contrast, Spi-C, a factor that is poorly studied in mammals, appears to have a unique spleen-specific expression pattern in skate. The expression of Spi-C in the spleen does not shift between embryo and adult, guaranteeing that at least one PU.1/Spi family member is present in each site of Igα cells in the skate. Although it is unknown exactly what role Spi-C plays in skate or mouse, its expression in the skate spleen does correlate well with the observation that Spi-C is specifically expressed in mature B cells but not at earlier B cell developmental stages in mammals (49).

By analogy with mammals, myeloid cells in the Leydig and epigonal organs (9) could be alternative sources of expression of factors such as PU.1 and Ikaros. High-level expression of Ikaros in all lymphoid sites is consistent with a continuing role in T and B cell development, in addition to possible uses in myeloid cells. Unlike Ikaros, PU.1 and Spi-D are only weakly expressed in the IgM−-rich spleen, raising the possibility that these two E26 (retrovirus) transformation-specific gene (Ets) family members are not significantly expressed in B cells at all, but are instead restricted primarily to myeloid cells, especially in the adult skate. Studies from another laboratory provide additional evidence that evolutionary change could have affected the roles of Ets family factors in lymphoid development during the vertebrate radiations. No cis-regulatory analysis has yet been reported for the B cell genes of the skate; however, the role for Ets family members in regulation of the channel catfish Ig H chain enhancer is drastically reduced as compared with mouse (50, 51). Thus, it is possible that teleosts and elasmobranchs make substantially less use of Spi-family Ets factors in mature B cells, generally, than do mammals.

Whereas some of the differences observed between mouse and skate transcription factor patterns may represent true divergence of regulation and/or function, other aspects of skate transcription factor usage that initially appear surprising may be representative of conserved features of hemopoietic development that have been unappreciated in the mammalian system heretofore. One such example may be the reciprocal expression of Runx2 and Runx3 in the myeloid/B cell-rich Leydig organ and the thymus, respectively. Although until recently neither a B/myeloid role for Runx2 nor a T cell role for Runx3 would have been predicted from known activities of their mammalian orthologs, Runx3 is in fact important in mouse thymocyte development (24, 25). Another example is the high expression of Helios in the adult skate spleen, which had not been previously reported but now appears to be conserved according to RT-PCR analyses in the mouse spleen (data not shown).

In summary, we show that, by the time of the divergence of the cartilaginous fish from the lineage leading to the mammals, the PU.1, Ikaros, GATA-1/2/3, Runx, Pax, and EBF-1/2/3 families had duplicated and diverged, and specific members had been recruited to expression domains involved in the control of lymphocyte development. Whereas the organ sites of hemopoiesis have shifted in phylogeny, the association between specific transcription factor combinations and lymphoid development remain strong. Such observations suggest that hemopoietic cell type specification must occur independently of broader embryonic patterning in all jawed vertebrates. This special property may relate to the migratory nature of the mature cells, and to the requirement that hemopoietic stem cells must generate differentiated progeny throughout the lifetime of the animal. Overall, these studies support the possibility that identification of hemopoietic cells expressing specific transcription factor combinations in lower deuterostomes can be identified even in animals that lack canonical lymphoid or myeloid organs.

Note added in proof. A recent paper by Schwitzer and DeKoter (52) shows that, even in the mouse, PU.1 and Spi-B are not strictly required for Ig gene expression.

Acknowledgments

We thank the Caltech DNA Sequencing Facility for technical assistance.

References


