

Evolution of TFs and Enhancers

Subjects: Immunology | Genetics & Heredity

Contributor: Masaki Miyazaki

Adaptive immunity relies on the V(D)J DNA recombination of immunoglobulin (Ig) and T cell receptor (TCR) genes, which enables the recognition of highly diverse antigens and the elicitation of antigen-specific immune responses. This process is mediated by recombination-activating gene (Rag) 1 and Rag2 (Rag1/2), whose expression is strictly controlled in a cell type-specific manner; the expression of Rag1/2 genes represents a hallmark of lymphoid lineage commitment. Although Rag genes are known to be evolutionally conserved among jawed vertebrates, how Rag genes are regulated by lineage-specific transcription factors (TFs) and how their regulatory system evolved among vertebrates have not been fully elucidated. Here, we review the current body of knowledge concerning the cis-regulatory elements (CREs) of Rag genes and the evolution of the basic helix-loop-helix TF E protein regulating Rag gene CREs, as well as the evolution of the antagonist of this protein, the Id protein. This may help to understand how the adaptive immune system develops along with the evolution of responsible TFs and enhancers.

Keywords: adaptive immune system ; Rag1/2 gene enhancers ; E protein ; Id protein ; bHLH transcription factors ; T and B cell development

1. Introduction

Our body is protected from invading pathogens by immune responses, which are primarily mediated by two distinct types of cells, the adaptive and innate immune cells. These cells cooperatively function to induce inflammatory responses to eliminate pathogens from the body. Adaptive immune cells, such as T and B cells, elicit pathogen-specific immune responses through the recognition of specific antigens, while innate immune cells, including macrophages, neutrophils, dendritic cells, and histiocytes, are activated by pattern recognition receptors (PRRs), which recognize distinct microbial components. Because the Rag1/2 genes are exclusively expressed in T cell and B cell progenitor/precursor stages, their expression implies adaptive lymphoid lineage commitment ^[1].

As well as T cell, upon B cell lineage commitment from common lymphoid progenitors (CLPs) in the bone marrow, CLPs give rise to adaptive lymphoid cells (T and B cell), innate lymphoid cells (ILCs), and plasmacytoid dendritic cells (pDCs). Notably, ILCs and T cells show functional similarities in cytokine production, and they commonly express Bcl11b, Tcf1, Gata3, and Runx during their development and activation ^[2]. In T cells, an anti-silencer element (ASE), which is located 73 kb upstream of the Rag2 gene and is 8 kb in length, is essential for Rag1/2 gene expression in developing T cells, but not in developing B cells ^[3].

2. Regulation of *Rag1/2* Gene by T or B Cell-Specific Enhancers

2.1. CREs for the Rag Gene, and Lineage-Specific Transcription Factors

The first wave of RAG expression is required for the recombination of TCR β and IgH genes in pro-T and pro-B cells, respectively. After the β -selection of pro-T cells and pre-BCR selection of pro-B cells, Rag expression is transiently downregulated during the developmental transition toward precursor stages (DP and pre-B cells). Both in mouse and human, impairment or loss of Rag gene expression and functions results in severe combined immunodeficiency, resulting from developmental arrest at pro-T and pro-B cell stages ^{[4][5]}. Deletion of Erag, which is 23 kb upstream of the Rag2 gene, caused impaired Rag1/2 expression in pro-B cells and a moderate developmental block at the pro-B stage but did not affect the Rag gene expression in T cell development ^[6].

Tcf1, Bcl11b, Gata3, Runx1, Satb1, and Ikaros; B cells: Pax5, Ebf1, Foxo1, Ets1, Irf4, and Ikaros) ^{[7][8][9][10]}. These results indicated that E-protein binding to the T cell-specific Rag gene enhancer is required for T cell-specific spatial interactions to enhance Rag1/2 expression ^{[11][7]}. Notably, blocking E2A binding to the Rag1 gene promoter region (R1pro) by generating E-box motif mutations alone resulted in the complete loss of Rag1 expression without affecting Rag2 expression in both developing T and B cells, leading to developmental arrest at the pro-T and pro-B cell stages ^[7]. Taken

together, these results strongly suggest that the activities of T cell-specific enhancer and Rag1 promoter depend on the binding of E2A to these regions and that E2A is a core TF that specifies the adaptive lymphoid cell identity through the regulation of Rag gene expression.

2.2. Evolution of Rag Gene Enhancer

Enhancer regions play a crucial role in precise pattern and amounts of gene expression during development, and divergence of the DNA sequence within enhancer region is considered to be related to the phenotypic variations among species [12]. This suggests that the phylogenetic conservation of DNA sequences within Rag gene enhancers reflect the evolution of Rag gene regulation. Thus, we investigated the conservation of R-TEn, R1B, and R2B regions and E-box motifs in these regions [7]. We found that DNA sequence similarities in *R-TEn* and *R2B* are readily observed among mammals, most birds, and reptiles; however, sequence similarities of these enhancers are not noticeable in the corresponding genomic regions of amphibians and fishes (Figure 1) [7]. Thus, we proposed that terrestrial animals evolutionarily acquired the E protein-mediated regulatory mechanisms as enhancers to increase the Rag gene expression, which induce higher expression of Rag genes and enable a diverse range of TCR and Ig gene recombination to protect our bodies from a wide range of pathogens.

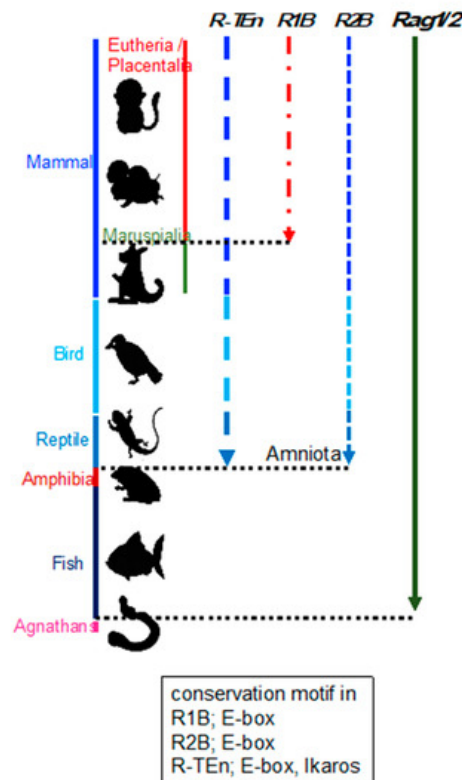


Figure 1. Schematic summary of the conservation of *R-TEn*, *R1B*, and *R2B* among vertebrates. Black, dotted lines indicate the border between placentalia and marsupialia, reptile and amphibia, and fish and agnathans. The conserved motifs in each enhancer region are shown in the box [7].

Regarding the evolution of AIS among vertebrates, cytidine deaminases CDA1 and CDA2 in jawless vertebrates are counterparts of Rag1 and Rag2 in jawed vertebrates and evolutionarily developed AIS as genome editors [13][14][15]. Furthermore, the recombination of Ig and TCR in fish seems to be more diverse than that in mammals, for example, the plasticity of T/B cells and the repertoire usage of TCR and Ig [16]. Given that the locations of B cell development among birds, reptiles, amphibians, and fish are different, it is reasonable that the variation in enhancer regions among species produces diversification of Rag1/2 gene regulation, such as timing. Considering this, it is surprising that both enhancer and promoter activities are critically controlled by E protein binding.

3. E Proteins and Id Proteins in Adaptive Lymphocyte Development

E proteins are basic helix-loop-helix (bHLH) transcription factors involved in multiple developmental processes. E proteins bind as homodimers or heterodimers to the E-box motif (CANNTG) within enhancer regions of their target genes. Id proteins contain an HLH domain missing the basic region that is essential for specific DNA binding and form heterodimers with bHLH proteins such as E proteins [17]. When the Id protein forms heterodimers with the E protein,

(Tcf3) is critically required for B cell lineage commitment [18][19] and the E2A gene encodes E12 and E47 proteins, which are generated by differential splicing [20][21]. In lymphoid progenitor cells, E2A orchestrates the B cell fate, along with Ebf1, Foxo1, and other TFs [22][23]. Upon T cell lineage commitment, E2A and HEB act in synergy to establish T cell identity and to suppress ILC development [24]. Likewise, HEB plays a role in iNKT cell development [25], and E2A and HEB also play important roles in the positive selection of DP thymocytes [26].

In B cell development, Id3 is induced in response to TGF β signaling for survival during early B cell development [27]. Id3 is highly expressed in naïve mature B cells and downregulated in activated germinal center B (GCB) cells, while E2A protein abundance is low in naïve B cells but high in GCB cells to induce AID expression in cooperation with E2-2 [28][29]. In T cell development, Id3 is first upregulated by pre-TCR signaling in DN3 cells and further upregulated upon positive selection of TCR signaling in DP cells [30][31]. Furthermore, Id3 plays a key role in follicular helper T (TFH) and follicular cytotoxic T(TFC) cell development through the regulation of CXCR5 expression [32][33][34].

4. Evolution of E and Id Proteins

In this section, we address the question of how the E–Emc is a negative feedback regulator that prevents runaway self-stimulation of Da gene expression in *Drosophila*. Coupled transcriptional feedback loops maintain the widespread Emc expression that restrains Da activity to induce neurons [35], suggesting that the transcriptional regulation system by E and Id proteins is conserved from the common ancestor of mammals and *Drosophila*.

Three E protein homologs and two Id protein homologs were found in the lamprey (*Petromyzon marinus*) (Figure 2). A reconstructed maximum likelihood phylogenetic tree of E protein homologs indicates that homologs of jawed vertebrates form three clades for E2A, E2-2, and HEB, strongly suggest that these paralogs were generated through the widely recognized two rounds of whole genome duplication (WGD) in vertebrates [36]. It is plausible that ancestral jawed vertebrates probably had four paralogs for each of the E

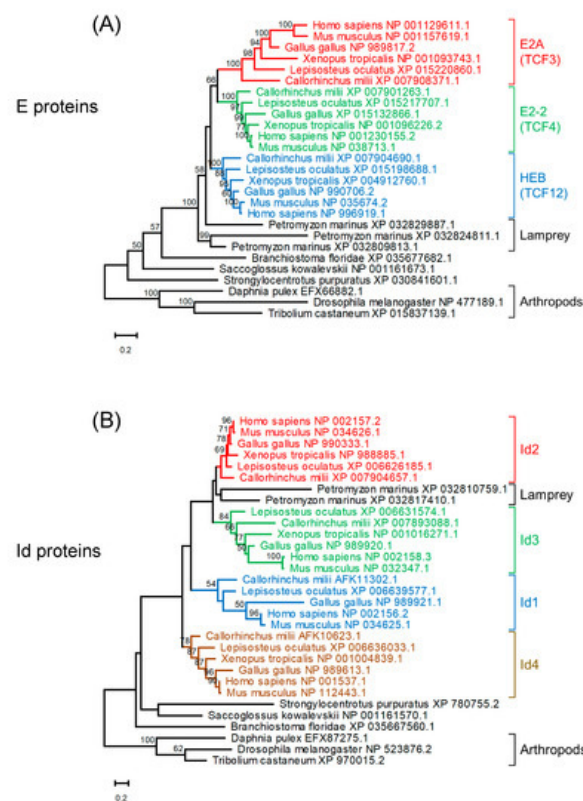


Figure 2. Maximum likelihood phylogenetic trees of homologs of E proteins (A) and Id proteins (B). Sequences were aligned using MAFFT (v7.453) [37] with default parameters. Tree reconstruction was performed using RAXML (version 8.2.12) [38] with the JTT + F substitution model and PROTGAMMA parameter with 100 bootstrap replicates. Phylogenetic trees were visualized using MEGA-X (version 10.2.4) [39]. Bootstrap values are given along the branches.

References

1. Kuo, T.C.; Schlissel, M.S. Mechanisms controlling expression of the RAG locus during lymphocyte development. *Curr. Opin. Immunol.* 2009, 21, 173–178.

2. Eberl, G.; Colonna, M.; Di Santo, J.P.; McKenzie, A.N.J. Innate lymphoid cells: A new paradigm in immunology. *Science* 2015, 348, aaa6566.
3. Yannoutsos, N.; Barreto, V.M.; Misulovin, Z.; Gazumyan, A.; Yu, W.; Rajewsky, N.; Peixoto, B.R.; Eisenreich, T.; Nussenzweig, M.C. A cis element in the recombination activating gene locus regulates gene expression by counteracting a distant silencer. *Nat. Immunol.* 2004, 5, 443–450.
4. Mombaerts, P.; Iacomini, J.; Johnson, R.S.; Herrup, K.; Tonegawa, S.; Papaioannou, V.E. RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* 1992, 68, 869–877.
5. Shinkai, Y.; Rathbun, G.; Lam, K.P.; Oltz, E.M.; Stewart, V.; Mendelsohn, M.; Charron, J.; Datta, M.; Young, F.; Stall, A.M.; et al. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* 1992, 68, 855–867.
6. Hsu, L.-Y.; Luring, J.; Liang, H.-E.; Greenbaum, S.; Cado, D.; Zhuang, Y.; Schlissel, M.S. A Conserved Transcriptional Enhancer Regulates RAG Gene Expression in Developing B Cells. *Immunity* 2003, 19, 105–117.
7. Miyazaki, K.; Watanabe, H.; Yoshikawa, G.; Chen, K.; Hidaka, R.; Aitani, Y.; Osawa, K.; Takeda, R.; Ochi, Y.; Tani-Ichi, S.; et al. The transcription factor E2A activates multiple enhancers that drive Rag expression in developing T and B cells. *Sci. Immunol.* 2020, 5, eabb1455.
8. Hao, B.; Naik, A.; Watanabe, A.; Tanaka, H.; Chen, L.; Richards, H.W.; Kondo, M.; Taniuchi, I.; Kohwi, Y.; Kohwi-Shigematsu, T.; et al. An anti-silencer– and SATB1-dependent chromatin hub regulates Rag1 and Rag2 gene expression during thymocyte development. *J. Exp. Med.* 2015, 212, 809–824.
9. Naik, A.; Byrd, A.T.; Lucander, A.C.; Krangel, M.S. Hierarchical assembly and disassembly of a transcriptionally active RAG locus in CD4+CD8+ thymocytes. *J. Exp. Med.* 2019, 216, 231–243.
10. Amin, R.H.; Schlissel, M.S. Foxo1 directly regulates the transcription of recombination-activating genes during B cell development. *Nat. Immunol.* 2008, 9, 613–622.
11. Miyazaki, K.; Miyazaki, M. The interplay between chromatin architecture and lineage-specific transcription factors and the regulation of Rag gene expression. *Front. Immunol.* 2021, 12, 6597612021.
12. Long, H.K.; Prescott, S.L.; Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* 2016, 167, 1170–1187.
13. Rogozin, I.B.; Iyer, L.M.; Liang, L.; Glazko, G.V.; Liston, V.G.; I Pavlov, Y.; Aravind, L.; Pancer, Z. Evolution and diversification of lamprey antigen receptors: Evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat. Immunol.* 2007, 8, 647–656.
14. Trancoso, I.; Morimoto, R.; Boehm, T. Co-evolution of mutagenic genome editors and vertebrate adaptive immunity. *Curr. Opin. Immunol.* 2020, 65, 32–41.
15. Morimoto, R.; O'Meara, C.P.; Holland, S.J.; Trancoso, I.; Souissi, A.; Schorpp, M.; Vassaux, D.; Iwanami, N.; Giorgetti, O.B.; Evanno, G.; et al. Cytidine deaminase 2 is required for VLRB antibody gene assembly in lampreys. *Sci. Immunol.* 2020, 5, eaba0925.
16. Hsu, E.; Pulham, N.; Rumfelt, L.L.; Flajnik, M.F. The plasticity of immunoglobulin gene systems in evolution. *Immunol. Rev.* 2006, 210, 8–26.
17. Kee, B.L. E and ID proteins branch out. *Nat. Rev. Immunol.* 2009, 9, 175–184.
18. Bain, G.; Maandag, E.C.; Izon, D.J.; Amsen, D.; Kruisbeek, A.M.; Weintraub, B.C.; Krop, I.; Schlissel, M.S.; Feeney, A.J.; Van Roon, M.; et al. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* 1994, 79, 885–892.
19. Zhuang, Y.; Soriano, P.; Weintraub, H. The helix-loop-helix gene E2A is required for B cell formation. *Cell* 1994, 79, 875–884.
20. Murre, C.; McCaw, P.S.; Baltimore, D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 1989, 56, 777–783.
21. Yamazaki, T.; Liu, L.; Lazarev, D.; Al-Zain, A.; Fomin, V.; Yeung, P.L.; Chambers, S.M.; Lu, C.-W.; Studer, L.; Manley, J.L. TCF3 alternative splicing controlled by hnRNP H/F regulates E-cadherin expression and hESC pluripotency. *Genes Dev.* 2018, 32, 1161–1174.
22. Lin, Y.C.; Jhunjhunwala, S.; Benner, C.; Heinz, S.; Welinder, E.; Mansson, R.; Sigvardsson, M.; Hagman, J.; Espinoza, C.; Dutkowski, J.; et al. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat. Immunol.* 2010, 11, 635–643.
23. Lin, Y.C.; Benner, C.; Mansson, R.; Heinz, S.; Miyazaki, K.; Miyazaki, M.; Chandra, V.; Bossen, C.; Glass, C.K.; Murre, C. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate

B cell fate. *Nat. Immunol.* 2012, 13, 1196–1204.

24. Miyazaki, M.; Miyazaki, K.; Chen, K.; Jin, Y.; Turner, J.; Moore, A.J.; Saito, R.; Yoshida, K.; Ogawa, S.; Rodewald, H.-R.; et al. The E-Id Protein Axis Specifies Adaptive Lymphoid Cell Identity and Suppresses Thymic Innate Lymphoid Cell Development. *Immunity* 2017, 46, 818–834.e4.
25. D'cruz, L.M.; Knell, J.; Fujimoto, J.K.; Goldrath, A.W. An essential role for the transcription factor HEB in thymocyte survival, Tcr α rearrangement and the development of natural killer T cells. *Nat. Immunol.* 2010, 11, 240–249.
26. Jones, M.E.; Zhuang, Y. Acquisition of a Functional T Cell Receptor during T Lymphocyte Development Is Enforced by HEB and E2A Transcription Factors. *Immunity* 2007, 27, 860–870.
27. Kee, B.L.; Rivera, R.R.; Murre, C. Id3 inhibits B lymphocyte progenitor growth and survival in response to TGF- β . *Nat. Immunol.* 2001, 2, 242–247.
28. Chen, S.; Miyazaki, M.; Chandra, V.; Fisch, K.M.; Chang, A.N.; Murre, C. Id3 Orchestrates Germinal Center B Cell Development. *Mol. Cell. Biol.* 2016, 36, 2543–2552.
29. Gloury, R.; Zotos, D.; Zuidschewoude, M.; Masson, F.; Liao, Y.; Hasbold, J.; Corcoran, L.M.; Hodgkin, P.D.; Belz, G.T.; Shi, W.; et al. Dynamic changes in Id3 and E-protein activity orchestrate germinal center and plasma cell development. *J. Exp. Med.* 2016, 213, 1095–1111.
30. Miyazaki, M.; Rivera, R.R.; Miyazaki, K.; Lin, Y.C.; Agata, Y.; Murre, C. The opposing roles of the transcription factor E2A and its antagonist Id3 that orchestrate and enforce the naive fate of T cells. *Nat. Immunol.* 2011, 12, 992–1001.
31. Engel, I.; Murre, C. E2A proteins enforce a proliferation checkpoint in developing thymocytes. *EMBO J.* 2003, 23, 202–211.
32. Miyazaki, M.; Miyazaki, K.; Chen, S.; Chandra, V.; Wagatsuma, K.; Agata, Y.; Rodewald, H.-R.; Saito, R.; Chang, A.N.; Varki, N. The E–Id protein axis modulates the activities of the PI3K–AKT–mTORC1–Hif1 α and c-myc/p19Arf pathways to suppress innate variant TFH cell development, thymocyte expansion, and lymphomagenesis. *Genes Dev.* 2015, 29, 409–425.
33. Liu, X.; Chen, X.; Zhong, B.; Wang, A.; Wang, X.; Chu, F.; Nurieva, R.I.; Yan, X.; Chen, P.; Van Der Flier, L.G.; et al. Transcription factor achaete-scute homologue 2 initiates follicular T-helper-cell development. *Nat. Cell Biol.* 2014, 507, 513–518.
34. Leong, Y.A.; Chen, Y.; Ong, H.S.; Wu, D.; Man, K.; Deleage, C.; Minnich, M.; Meckiff, B.J.; Wei, Y.; Hou, Z.; et al. CXCR5 $^{+}$ follicular cytotoxic T cells control viral infection in B cell follicles. *Nat. Immunol.* 2016, 17, 1187–1196.
35. Bhattacharya, A.; Baker, N.E. A Network of Broadly Expressed HLH Genes Regulates Tissue-Specific Cell Fates. *Cell* 2011, 147, 881–892.
36. Simakov, O.; Marlétaz, F.; Yue, J.-X.; O'Connell, B.; Jenkins, J.; Brandt, A.; Calef, R.; Tung, C.-H.; Huang, T.-K.; Schmutz, J.; et al. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol. Evol.* 2020, 4, 820–830.
37. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 2013, 30, 772–780.
38. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014, 30, 1312–1313.
39. Kumar, S.; Stecher, G.; Li, M.; Niyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 2018, 35, 1547–1549.