The concerted action of E2-2 and HEB is critical for early lymphoid specification

Thibault Bouderlique[1#], Lucia Peña Perez[1#], Shabnam Kharazi[1], Miriam Hils[2], Xiaoze Li[1], Aleksandra Krstic[1], Ayla De Paepe[1], Christian Schachtrup[3], Charlotte Gustafsson[1], Dan Holmberg[4], Kristina Schachtrup[2] and Robert Månsson[1,5*].

[1] Center for Hematology and Regenerative Medicine, Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden

[2] Center for Chronic Immunodeficiency (CCI), Medical Center, University of Freiburg, Faculty of Medicine & Faculty of Biology, Freiburg, Germany

[3] Institute of Anatomy and Cell Biology, University of Freiburg, Faculty of Medicine, Freiburg, Germany

[4] Lund University Diabetes Center, Lund University, Malmö, Sweden

[5] Hematology Center, Karolinska University Hospital, Stockholm, Sweden

[#] Contributed equally to this work

[*] Correspondence should be addressed to RM (robert.mansson@ki.se)

Abstract

The apparition of adaptive immunity in *Gnathostomata* correlates with the expansion of the E-protein family to encompass E2-2, HEB and E2A. Within the family, E2-2 and HEB are more closely evolutionarily related but their concerted action in hematopoiesis remains to be explored. Here we show that the combined disruption of E2-2 and HEB results in failure to express the early lymphoid program in CLPs and a near complete block in B-cell development. In the thymus, ETPs were reduced and T-cell development perturbed, resulting in reduced CD4 T- and increased γδ T-cell numbers. In contrast, HSCs, erythro-myeloid progenitors and innate immune cells were unaffected showing that E2-2 and HEB are dispensable for the ancestral hematopoietic lineages. Taken together, this E-protein dependence suggests that the appearance of the full *Gnathostomata* E-protein repertoire was critical to reinforce the gene regulatory circuits that drove the emergence and expansion of the lineages constituting humoral immunity.

Large *Eumetazoan* rely on an efficient system of innate and adaptive immune cells to survive and reach reproductive age [1-3]. The different cells of the hematopoietic system are all generated from hematopoietic stem cells (HSCs) [4]. Lymphoid specification is initiated in lymphoid primed multipotent progenitors (LMPPs) that start to express genes associated with adaptive immune cells [5,6]. LMPPs subsequently give rise to common lymphoid precursors (CLP) [7]. Within the heterogeneous CLP population, the LY6D$^+$ fraction is further specified towards a B-lineage fate [8,9] and contains the first B-lineage committed cells that subsequently give rise to mature B-cells [10,11]. Early lymphoid precursors leave the bone marrow to seed the thymus where they further develop into early T-cell progenitors (ETP) that give rise to mature T-cells [12]. Similarly, the innate immune cells develop from different progenitors within the myeloid branch, developing from granulocyte-macrophage progenitors (GMPs) [13], while natural killer (NK) cells and part of the dendritic cells (DC) develop from the CLP [14,15].

The origin of the *Gnathostomata* (jawed vertebrate) hematopoietic system can be traced far back in evolutionary history with phagocytic and cytotoxic innate immune cells being found across the *Bilateria* [16] and the erythroid/megakaryocyte lineages appearing in the *Agnatha* [17]. Similarly, lymphoid-like cells are present in the *Agnatha* [18], *Urochordata* [19] and *Cephalochordata* [20]. However, while genes intimately associated with adaptive immunity - including RAG [21,22], histocompatibility genes [23,24] and immune type receptors [23,25,26] - are found in lower *Deuterostomata*, B- and T-cells mediated adaptive immunity emerged only in the *Gnathostomata*. The appearance of new transcription factor (TF) genes drive the apparition of novel cell types [27]. The appearance of adaptive immunity in the *Gnathostomata* correlates with a dramatic increase in TF genes [1,28]. As part of this expansion, the full *Gnathostomata* basic helix-loop-helix E-protein family [29,30] consisting of E2A (Tcf3), HEB (Tcf12) and E2-2 (Tcf4) emerged.

It has been proposed, that E2A is more closely related to the ancestral E-proteins while E2-2 and HEB are less evolutionarily conserved and display expression patterns more restricted to vertebrate-specific structures [30,31]. This suggests that E2A should govern ancestral functions while HEB and E2-2 should govern novel functions that emerged concomitantly to the rise of the *Gnathostomata*. In line with this, E2A is the only E-protein reported to impact HSC function and the development of the myeloid- and erythro/megakaryocytic lineages [32,33]. In contrast, all the E-proteins promote development of B- and T-cells [9,33-43]. The potential role of E2-2 in stem- and progenitor cells remains largely unexplored.

Here we confirm that E2-2 and HEB are evolutionary related and we found that their coordinated action is critical for the development of early lymphoid progenitors. Mice lacking both E2-2 and HEB display an almost complete block in B-cell development at the level of the CLP and the few generated immature B-cells preferentially develop into marginal zone (MZ) B-cells. Similarly, we found T-cell development to be perturbed, resulting in reduced numbers of CD4 T-cells and increased numbers of γδ T-cells. In contrast, HSCs, erythro-myeloid development and the generation of innate immune cells were unperturbed. Together, this suggests that E2-2 and HEB are dispensable for ancestral hematopoietic lineages and that the appearance of the full *Gnathostomata* E-protein repertoire promoted the apparition of humoral immunity.

Results

E2-2 and HEB are evolutionarily related

To investigate the evolutionary history of the E-proteins in *Bilateria* in light of recent sequencing data, we analyzed similarities between cDNA and amino acid sequences of the E-proteins across *Animalia* (Fig. 1A). The E-proteins of *Protostomata* and non-gnathostome *Deuterostomata* clustered together (Fig. 1B) and displayed relatively high sequence divergence

(Fig. 1C) on the cDNA level. The *Gnathostomata* E-proteins, in contrast, formed a separate clade (Fig. 1B) and displayed comparably higher similarity (Fig. 1C). Within the *Gnathostomata* clade, E2-2 and HEB formed a separate branch from that of E2A (Fig. 1C). Similar results were obtained from the analysis of the amino acid sequences (Fig. S1). While only providing weak support for E2A being more closely related to the ancestral E-protein, it confirms a closer evolutionary relation between E2-2 and HEB [30]. It could hence be hypothesized that E2-2 and HEB are functionally related and together support the development of cell lineages specific to the jawed vertebrate hematopoietic system.

Deletion of E2-2 and HEB does not impact HSCs and erythro-myeloid progenitors

To investigate the role of E2-2 and HEB in hematopoiesis, we used the Vav^{iCre} mouse strain to mediate conditional deletion of E2-2 ($E2\text{-}2^{f/f}Vav^{iCre}$) or E2-2 together with HEB ($E2\text{-}2^{f/f}HEB^{f/f}Vav^{iCre}$) in the hematopoietic system. Cre mediated deletion of the floxed exons was verified in RNAseq data from FACS sorted cells (Fig. S2A-C). We observed a difference in bone marrow (BM) cellularity only between littermate control mice (lacking Vav^{iCre}) and $E2\text{-}2^{f/f}HEB^{f/f}Vav^{iCre}$mice (Fig. 1D). However, no differences were observed in the number of HSCs (Fig. 2E, S3A), megakaryocyte, erythroid and myeloid progenitors (Fig. 2F, S3B). While being expressed in the HSC and erythro-myeloid progenitors (Fig. S2D), E2-2 and HEB are hence dispensable for the maintenance and generation of these cell types in steady state [33]. This leaves E2A as the sole E-protein needed for HSCs and erythro-myeloid development [32].

Generation of innate immune cells except pDCs are unaffected by E2-2 and HEB deletion

While the total number of spleen cells was reduced in $E2\text{-}2^{f/f}HEB^{f/f}Vav^{iCre}$mice (Fig. 1G), no effect was observed on the number of mature myeloid cells (Fig. 1H and S3C). This further support that E2-2 and HEB are dispensable for myelopoiesis. Similarly, natural killer cell

numbers were unaffected (Fig. 1H and S3C). In contrast, plasmacytoid dendritic cells (pDC) were reduced by 86% by the deletion of E2-2 and by 98% with the additional removal of HEB (Fig. 1I-J) indicating a previously unrecognized dependence on HEB for pDC development [44].

E2-2 and HEB are needed for proper generation of B-cells and CD4 T-cells.

As the total number of spleen cells was reduced (Fig. 1G), we next investigated the effect of deletion of E2-2 and HEB on the adaptive immune cells. The number of T-cells in the spleen was not impacted by the depletion of E2-2, but E2-2$^{f/f}$HEB$^{f/f}$Vav-iCre mice displayed a 60% reduction in CD4 T-cells (Fig. 2A). The total number of B-cells in spleen was reduced by 42% by the deletion of E2-2 and by 68% when HEB was additionally depleted (Fig. 2B). This reduction reflected a sharp drop in transitional- and follicular B-cells (FoB) while marginal zone (MZ) B-cell numbers remained unaffected in E2-2$^{f/f}$Vav-iCre and increased in E2-2$^{f/f}$HEB$^{f/f}$Vav-iCre mice (Fig. 2C-E). This suggests that the removal of E2-2 and HEB heavily promoted the generation of MZ from transitional B-cells at the expense of FoB [45,46]. Taken together, these data show that cell lineages with ancestral functions (erythro-myeloid and cytotoxic cells)[16,17,47-50] are independent of E2-2 and HEB. In contrast, cell lineages central to adaptive (humoral) immunity (B- and CD4 T-cells) or at the interphase between innate and adaptive immunity (pDC) are dependent on E2-2 and HEB.

The development of CD4 and γδ T-cells is perturbed by the loss of E2-2 and HEB.

With E2-2$^{f/f}$HEB$^{f/f}$Vav^{iCre} mice displaying decreased CD4 T-cells in the periphery (Fig. 2A), we next investigated T-cell development in the thymus (Fig. 3A). Depletion of E2-2 had no effect on thymic cellularity (Fig. 3B). However, E2-2$^{f/f}$Vav^{iCre} mice displayed a visible decrease in ETPs and DN2 followed by decreases in DN3E and IS(8)P (Fig. 3D). In line with previous studies of HEB knock-out animals [34], the additional deletion of HEB perturbed T-cell

development (Fig. 3A) with decreased total cellularity in thymus (Fig. 3B); significant decreases in ETP and DN2 (Fig. 3C); increases in DN3E, DN3L and IS(8)P (Fig. 3D); but markedly reduced DPs (Fig. 3D). In spite of the marked reduction in DPs, CD8 T-cells were present in normal numbers (Fig. 3E) while CD4 T-cells were severely reduced (>85%). Additionally, E2-2$^{f/f}$HEB$^{f/f}$Vav^{iCre} animals displayed markedly increased (6-fold) γδ T-cells numbers (Fig. 3E) [34,51].

Looking closer at the progressive generation of cells from the prior stage in each developmental transition in E2-2$^{f/f}$HEB$^{f/f}$Vav^{iCre} mice, the altered cell numbers observed were mirrored by increased generation of DN3E, DN3L and IS(8)P before a sharp drop in the generation of DP TCRβ^{-}cells (Fig. 3F). While few in numbers, DP TCRβ^{+} were generated in increased numbers with CD4 and CD8 T-cells subsequently being generated at a normal and increased ratio respectively (Fig. 3F). Taken together these results demonstrate that E2-2 and HEB are critical for thymopoiesis and the normal generation of γδ and CD4 T-cells.

E2-2 and HEB are critical for lymphoid specification and B-cell development

The development of LMPPs and CLPs constitute the first steps of lymphoid specification [5,7,11]. Depletion of E2-2 alone did not significantly affect the number of LMPPs and Ly6D^{-}CLPs (Fig. 4A-C). However, the number of B-cell specified LY6D^{+}CLPs [8,11] was reduced by half (Fig. 4C). Correspondingly, total B-cells were reduced by 43% (Fig. 4D) with each stage in B-cell development displaying a 30-50% reduction (Fig. 4E-F). The additional loss of HEB, in contrast, lead to a 70% decrease in the number of LMPPs and LY6D^{-}CLP (Fig. 4A-C). In addition, the E2-2$^{f/f}$HEB$^{f/f}$Vav^{iCre} mice strikingly displayed a near complete loss of LY6D^{+}CLP (>98% reduction) (Fig. 4A-B). This phenotype is similar to what has previously been reported for E2A knock-out [33]. Accordingly, B-cell numbers were severely reduced (Fig. 4D) with a

>99% reduction in cell-numbers of maturing B-cells (Fig. 4E-F) and a 94% reduction in mature B-cells (Fig. 4E-F) in BM.

To analyze more closely the impact of E2-2 and HEB depletion on developmental transitions, we plotted cell number ratios at each consecutive stage of development as a fraction of the prior stage (Fig. 4G). E2-2 deficient mice only displayed a modest decrease in the generation of LY6D^{+}CLP from LY6D^{-}CLP while the consecutive generation of CD19^{+} developmental stages remained unaffected (Fig. 4G). This indicates that E2-2, similarly to what has been reported for HEB deficient animals [33], is important for the LY6D^{-} to LY6D^{+} CLP transition but largely dispensable for BM B-cell maturation. This is in line with the expression data showing that early lymphoid progenitors (LMPPs and CLPs) expressed similar levels of E2-2, HEB and E2A while HEB and E2A were the predominantly expressed E-proteins in B-lineage cells (Fig. S2D). The combined deletion of E2-2 and HEB severely impacted the generation of LMPPs, LY6D^{+}CLP and pre-B while, at the same time, seemingly increasing the generation of immature B-cells (Fig. 4G). Interestingly, the generation of proB cells from the few remaining LY6D^{+}CLP was unaffected (Fig. 4G), indicating that B-lineage commitment at this stage is unperturbed by the lack of E2-2 and HEB. Together, this demonstrates that the collaboration of E2-2 and HEB is critical for the generation of early lymphoid progenitors and the development of B-cells in the BM.

Expression of the early lymphoid program is disrupted by removal of E2-2 and HEB.

To better understand the mechanisms behind the impaired generation of the B-cell specified LY6D^{+}CLPs and to understand how it relates to the similar phenotype observed in mice lacking E2A [33], we characterized the transcriptional profiles of LY6D^{-}CLPs remaining in E2-2$^{f/f}$Vav^{iCre}, E2-2$^{f/f}$HEB$^{f/f}$Vav^{iCre} and E2A$^{f/f}$Vav^{iCre} mice using RNAseq (see Table S2 for sample information). Principal component analysis (PCA) of full expression profiles showed that

LY6D$^-$CLP cells from the different genotypes form a distinct cluster and hence represent the same population of cells regardless of disrupted E-protein gene(s) (Fig. 5A). A total of 150 genes displayed significant (corrected p-value <0.01) expression changes in LY6D$^-$CLPs in one or more of the analyzed strains (Fig. 5B). As expected, B-lineage related genes (including Ebf1, Blnk, Blk, Ets1, Dntt, Notch1, Rag1 and Rag2) were severely affected (Fig. 5B) and overall the gene set was functionally associated with lymphocyte differentiation and signaling (Fig. S4). While the majority of genes did not show highly significant changes in all genotypes (Fig. 5C-D), the overall pattern of the expression changes was highly similar in E2-2^{ff}HEB^{ff}Vav^{iCre} and E2A^{ff}Vav^{iCre} mice (Fig. 5B and D-E). The modest expression changes observed in the E2-2^{ff}Vav^{iCre} mice were often concordant to those observed in E2A^{ff}Vav^{iCre} mice and E2-2^{ff}HEB^{ff}Vav^{iCre} mice (Fig. 5B and D-E). Taken together, this shows that E2-2, HEB and E2A largely reinforce the same gene network in early lymphoid progenitors while having variable impact on the expression of individual genes.

E-proteins display partly overlapping association with chromatin

To further understand the mechanisms through which E2-2 and HEB control B lymphopoiesis, we analyzed the binding pattern of E2A, HEB and E2-2 in pro B-cells using ChIP-seq (see Table S2 for sample information). We identified 16510 E2A, 2167 HEB and 139 E2-2 high quality peaks (peak score ≥10) (Fig. 6A and S5A). E2-2 is the lowest expressed E-protein in pro B-cells (Fig. S2D) and the relatively low enrichment of E2-2 (most peaks have a peak score <10, Fig. S5A) limits accurate peak calling. Hence, E2-2 binding is likely underestimated. Peaks from the three E-protein ChIPseq experiments displayed highly significant enrichment of E-protein DNA binding motifs (Fig. S5B).

E2A mostly bound regions without identified HEB and E2-2 peaks (Fig. 6A and S5C). However, the majority of E2-2 and HEB peaks overlapped with those of E2A (Fig. 6A and

S5C). Most binding sites identified where localized in intronic and intergenic regions (Fig. 6B). In contrast to the other E-proteins, a significant fraction of E2A peaks were also found in promoter regions (Fig. 6B). No difference was observed in the E-protein DNA binding motifs of HEB and/or E2A, with the most enriched motifs containing the same bHLH motif core (CAGCTG) [52]. The most enriched motif in E2A peaks was present amongst the motifs found in the peaks common to E2A and HEB (Fig. S5D). Genes central to B-lineage, including IgH, CD19, Pax5, CD79a, CD79b, Igll1 and VpreB, were bound by the three E-proteins (Fig. 6C). A similar pattern was found for genes that are part of the E-protein-dependent early lymphoid program in LY6D^{-} CLPs (Fig. 5B), including Ebf1, Rag1, Rag2, Blnk, Dntt, Notch1 and Blk (Fig. 6D-E and S5E). These regions at the center of the lymphoid and B-lineage programs, bound by E2A and HEB, presented a higher enrichment for Ebf and Pax DNA binding motifs compared to regions with only detectable E2A peaks (Fig. S5F).

Taken together this further supports the notion that E2-2 and HEB critically reinforce the action of E2A on gene regulatory circuits critical to early lymphoid specification and B-cell development.

Discussion

Phylogenetic analysis of the E-proteins suggests that the *Gnathostomata* E-protein family arose through two subsequent duplications [30,53]. The initial duplication generated two E-protein loci that diverged into the proto-E2-2/HEB and E2A loci. The second duplication subsequently gave rise to the E2-2 and HEB loci as well as two E2A loci out of which one was eventually lost [53,54]. Being evolutionarily related, E2-2 and HEB could hence support similar functions. The E-proteins play critical roles in all the hematopoietic lineages [9,32-34,36,38-46,55,56]. However, the function of E2-2 in early hematopoietic development and its concerted action with the evolutionarily related E-protein HEB has not been thoroughly addressed.

The B-cell developmental pathway has been shown to critically rely on E2A and HEB, both TFs needed for the generation of LY6D^{+}CLPs [33]. We similarly found that the loss of E2-2 impaired the generation of LY6D^{+}CLPs while earlier progenitors (including LMPPs and LY6D^{-}CLPs) were unaffected. Further development of B-cells from LY6D^{+}CLPs was unperturbed with the subsequent stages appearing at expected ratios. This indicates that the reduced number of B-cells is primarily caused by the impaired LY6D^{-} to LY6D^{+} transition within the CLP compartment.

Interestingly, the combined deletion of E2-2 and HEB had a direct additive effect resulting in a near complete developmental block at the LY6D^{-}CLP stage and dramatically reduced B-cell numbers. Hence, E2-2 and HEB together have a critical and previously unrecognized role in supporting early lymphoid development in the BM. This phenotype is reminiscent of the one observed in E2A knockout animals [33]. In line with this, the genes of the early lymphoid program were associated with combined E2A and HEB binding. This further supports the notion that all three E-proteins, to a large extent, reinforce the same gene regulatory circuit in CLPs [33].

The dramatic reduction of BM B-lymphopoiesis in E2-2^{ff}Vav^{iCre} and E2-2^{ff}HEB^{ff}Vav^{iCre} animals was mirrored in reduced numbers of transitional (T1 and T2) B-cells and reduced FoBs. In contrast MZ B-cell numbers were maintained in E2-2^{ff}Vav^{iCre} mice or even expanded in E2-2^{ff}HEB^{ff}Vav^{iCre}mice. This means that MZ B-cells are generated at the expense of FoBs, confirming that the levels of all three E-proteins are critical to maintain a normal MZ to FoB ratio [45,46].

E2-2^{ff}Vav^{iCre}and E2-2^{ff}HEB^{ff}Vav^{iCre} mice displayed, on average, a 60% reduction in ETPs. Hence, while at a reduced level, E2A alone is sufficient to maintain thymic seeding in the absence of the other E-proteins [36,42]. T-cell development downstream of the ETP was significantly perturbed only in E2-2^{ff}HEB^{ff}Vav^{iCre} animals resulting in reduced CD4 T-cells

and increased generation of γδ T-cells. This phenotype is similar to the one observed in the HEB KO mice [34]. Hence, this suggests that E2-2 has limited impact on adult thymopoiesis [44].

In contrast to the clear impact of E2-2 and HEB on cells involved in humoral immunity (B- and CD4 T-cells), the loss of these TFs did not affect HSC numbers, erythro-myeloid progenitors nor the production of the major innate immune cells lineages including granulocytes, macrophages and natural killer cells. Similarly the cytotoxic branch of adaptive immunity (CD8 T-cells) was unaffected. Taken together, this suggests that E2-2 and HEB - in contrast to E2A [32,33] - are dispensable for lineages with ancestral functions. Functionally, this would suggest that E2A, while crucial for hematopoietic ancestral functions, was co-opted for lymphoid development and B-cell development.

Until recently, it was hypothesized that the adaptive immune system arose from an evolutionary "big bang" at the speciation of the *Gnathostomata* [1]. However, advances in genome sequencing of lower *Deuterostomata* has shifted this dogma by describing the presence of adaptive immunity related genes (including RAG, histocompatibility genes and immune type receptors) previously thought to be restricted to the *Gnathostomata* [21-26,48]. The presence of lymphoid-like cells in lower chordates [18-20] further suggests that the lymphoid genetic toolbox was present before the emergence of humoral immunity. Accordingly, the E-protein dependence of the early lymphoid program suggest that the appearance of the full *Gnathostomata* E-protein repertoire was crucial for reinforcing the gene regulatory circuits that drove the emergence and expansion of the hematopoietic lineages constituting humoral immunity.

Material and Methods

Animal studies

To generate mice lacking specific E-proteins in the hematopoietic system, Vav-iCre [57] was used in combination with conditional (floxed) E2-2 [58], HEB [41] and E2A [59] alleles. Mice were maintained on a C57BL/6 background and analyzed at 8 to 14 weeks of age. Animal studies were approved by the local ethics committee (ethical approval number S16-15).

Preparation of cells and flow cytometry

Bones, spleen and thymus were dissected, crushed in PBS with 2% FCS and cells were collected after passing through a 70 μm filter. They were then Fc-blocked (CD16/32; 93) and stained with combinations of the antibodies Sca1 (D7), CD105 (MJ7/18), CD41 (MWReg30), CD48 (HM48-1), CD3 (145-2C11), CD4 (RM4-5), CD8 (53-6.7), B220 (RA3-6B2), NK1.1 (PK136), Mac1 (M1/70), Gr1 (RB6-8C5), TER119 (TER-119), CD150 (TCF15-12F12.2), CD117 (2B8, eBioscience), CD127 (A7R34), CD44 (IM7), CD25 (PC61.5, eBioscience), CD19 (1D3, eBioscience), TcRβ (H57-597, eBioscience), TcRγδ (GL3, eBioscience), Ly6C (AL-21), Ly6G (1A8), MHCII (M5/114.15.2), CD11c (N418), PDCA1 (927), Ly6D (49H4), Flt3 (A2F10), IgD (11-26c.2a) and IgM (11/41, eBioscience). All antibodies were purchased from BD Biosciences unless otherwise indicated. For hematopoietic stem and progenitor cell (HSPC) isolation, the cells were subjected to lineage depletion using Dynabeads sheep anti rat IgG (Life Technologies) together with TER119, CD19, CD3, Gr1 and CD11b antibodies. Analysis and cell sorting was performed primarily on an LSR Fortessa and FACSAria IIu (BD Biosciences). Analysis of data was done using the Flowjo 9.9.6 software (Flowjo).

Phylogenetic analysis

The cDNA and amino acids sequences of the E-proteins from analyzed organisms were obtained through the E-ensembl repository [60]. See Table S1 for the sequences used in this study. Phylogenetic trees were constructed with MEGA7 [61] selecting the Maximum Likelihood

method based on the Tamura-Nei model; creating initial tree(s) using the Neighbor-Joining and BioNJ algorithms; and using a Gamma distribution with invariant sites. To assess the support of each node, the tree was bootstrapped 500 times.

RNA sequencing and analysis

5-10 x 10^3 cells were FACS-sorted into buffer RLT with β-mercaptoethanol and total RNA extracted using RNeasy Micro Kit (Qiagen, Hilden, Germany) with on-column DNase I treatment. Strand specific RNAseq libraries were prepared using the TotalScript™ RNA-seq kit (Epicentre, Madison, WI) together with custom made Tn5 (transposase). Barcoded libraries were pooled and pair-end sequenced (2x50 cycles) using the Illumina platform (NextSeq500, Illumina, San Diego, CA).

RNAseq reads were mapped using STAR [62], reads in exons quantified using HOMER [63] and significant changes identified using EdgeR. PCA analysis and display was performed using R (v3.3.3). For details see Supplemental Materials and Methods.

Pro B cell expansion cultures and fixation

$B220^+$ cells were isolated from bone marrow of ER-Cre mice [64] using magnetic cell separation (Miltenyi Biotec), expanded for six days in the presence of IL-7 and SCF to obtain pro B-cells. Pro B-cells were subsequently retrovirally transduced with Bcl-2, expanded for seven additional days with 5μM 4-Hydroxytamoxifen in the medium during the last 72 hours. Pro B-cells were fixed using EGS (1.5mM for 30min) in combination with PFA (1% for 10min) and stored as pellets in -80°C.

ChIP sequencing and analysis

Fixed pro B-cells were thawed, resuspended in SDS lysis buffer, sonicated, Triton X-100 was added and lysates subjected to ChIP by adding dynabeads pre-coupled with either antibodies against E2A, HEB or E2-2. ChIPed chromatin was washed and libraries prepared using reverse crosslinking in conjunction with adapter ligation or via direct amplification of tagmented libraries without prior DNA purification. For details see Supplemental Materials and Methods. Libraries were sequenced using the HiSeq2000 or NextSeq500 platforms (Illumina).

ChIPseq reads were mapped using bowtie2 [65]. Identification of peaks, peak overlaps and motif enrichment/identification analysis was done using HOMER's findPeaks, mergePeaks and findMotifsGenome.pl respectively [63]. Visualization was done using the UCSC genome browser. For details see Supplemental Materials and Methods.

Data availability

Raw sequencing data are available from the European Nucleotide Archive (ENA) under accession number PRJEB29568 (E-protein knock-out RNAseq and E-protein ChIPseq) and PRJEB20316 (wild-type C57bl/6 RNAseq data).

References

1 Flajnik, M. F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet* **11**, 47-59, doi:http://www.nature.com/nrg/journal/v11/n1/suppinfo/nrg2703_S1.html (2010).

2 Litman, G. W., Rast, J. P. & Fugmann, S. D. The origins of vertebrate adaptive immunity. *Nat Rev Immunol* **10**, 543-553 (2010).

3 Parra, D., Takizawa, F. & Sunyer, J. O. Evolution of B Cell Immunity. *Annual Review of Animal Biosciences* **1**, 65-97, doi:10.1146/annurev-animal-031412-103651 (2013).

4 Azevedo Portilho, N., Kobayashi, M. & Yoshimoto, M. What do the lineage tracing studies tell us? Consideration for hematopoietic stem cell origin, dynamics, and leukemia-initiating cells. *International Journal of Hematology*, doi:10.1007/s12185-018-2537-9 (2018).

5 Adolfsson, J. *et al.* Identification of Flt3+ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential: A Revised Road Map for Adult Blood Lineage Commitment. *Cell* **121**, 295-306, doi:https://doi.org/10.1016/j.cell.2005.02.013 (2005).

6 Månsson, R. *et al.* Molecular Evidence for Hierarchical Transcriptional Lineage Priming in Fetal and Adult Stem Cells and Multipotent Progenitors. *Immunity* **26**, 407-419, doi:https://doi.org/10.1016/j.immuni.2007.02.013 (2007).
7 Kondo, M., Weissman, I. L. & Akashi, K. Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow. *Cell* **91**, 661-672, doi:https://doi.org/10.1016/S0092-8674(00)80453-5 (1997).
8 Inlay, M. A. *et al.* Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes & Development* **23**, 2376-2381, doi:10.1101/gad.1836009 (2009).
9 Lin, Y. C. *et al.* A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat Immunol* **11**, 635-643, doi:http://www.nature.com/ni/journal/v11/n7/suppinfo/ni.1891_S1.html (2010).
10 Mansson, R. *et al.* Positive intergenic feedback circuitry, involving EBF1 and FOXO1, orchestrates B-cell fate. *Proceedings of the National Academy of Sciences* **109**, 21028-21033, doi:10.1073/pnas.1211427109 (2012).
11 Mansson, R. *et al.* Single-cell analysis of the common lymphoid progenitor compartment reveals functional and molecular heterogeneity. *Blood* **115**, 2601-2609, doi:10.1182/blood-2009-08-236398 (2010).
12 Zlotoff, D. A. & Bhandoola, A. Hematopoietic progenitor migration to the adult thymus. *Annals of the New York Academy of Sciences* **1217**, 122-138, doi:doi:10.1111/j.1749-6632.2010.05881.x (2011).
13 Marr, C., Strasser, M., Schwarzfischer, M., Schroeder, T. & Theis, F. J. Multi-scale modeling of GMP differentiation based on single-cell genealogies. *The FEBS Journal* **279**, 3488-3500, doi:doi:10.1111/j.1742-4658.2012.08664.x (2012).
14 Kondo, M. *et al.* Cell-fate conversion of lymphoid-committed progenitors by instructive actions of cytokines. *Nature* **407**, 383, doi:10.1038/35030112 https://http://www.nature.com/articles/35030112 - supplementary-information (2000).
15 Manz, M. G., Traver, D., Miyamoto, T., Weissman, I. L. & Akashi, K. Dendritic cell potentials of early lymphoid and myeloid progenitors. *Blood* **97**, 3333-3341, doi:10.1182/blood.V97.11.3333 (2001).
16 Nappi, A. J. & Ottaviani, E. Cytotoxicity and cytotoxic molecules in invertebrates. *BioEssays* **22**, 469-480, doi:10.1002/(SICI)1521-1878(200005)22:5<469::AID-BIES9>3.0.CO;2-4 (2000).
17 Svoboda, O. & Bartunek, P. Origins of the Vertebrate Erythro/Megakaryocytic System. *BioMed Research International* **2015**, 10, doi:10.1155/2015/632171 (2015).
18 Suzuki, T., Shin-I, T., Kohara, Y. & Kasahara, M. Transcriptome analysis of hagfish leukocytes: a framework for understanding the immune system of jawless fishes. *Developmental & Comparative Immunology* **28**, 993-1003, doi:http://dx.doi.org/10.1016/j.dci.2004.04.005 (2004).
19 Peddie, C. M. & Smith, V. J. 'Lymphocyte-like'cells in ascidians: Precursors for vertebrate lymphocytes? *Fish & Shellfish Immunology* **5**, 613-629, doi:http://dx.doi.org/10.1016/S1050-4648(95)80045-X (1995).
20 Huang, G. *et al.* The Identification of Lymphocyte-Like Cells and Lymphoid-Related Genes in Amphioxus Indicates the Twilight for the Emergency of Adaptive Immune System. *PLOS ONE* **2**, e206, doi:10.1371/journal.pone.0000206 (2007).
21 Kapitonov, V. V. & Koonin, E. V. Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon. *Biology Direct* **10**, 20, doi:10.1186/s13062-015-0055-8 (2015).

22 Morales Poole, J. R., Huang, S. F., Xu, A., Bayet, J. & Pontarotti, P. The RAG transposon is active through the deuterostome evolution and domesticated in jawed vertebrates. *Immunogenetics* **69**, 391-400, doi:10.1007/s00251-017-0979-5 (2017).
23 Azumi, K. *et al.* Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". *Immunogenetics* **55**, 570-581, doi:10.1007/s00251-003-0606-5 (2003).
24 Voskoboynik, A. *et al.* Identification of a Colonial Chordate Histocompatibility Gene. *Science* **341**, 384-387, doi:10.1126/science.1238036 (2013).
25 Du Pasquier, L., Zucchetti, I. & De Santis, R. Immunoglobulin superfamily receptors in protochordates: before RAG time. *Immunological Reviews* **198**, 233-248, doi:10.1111/j.0105-2896.2004.00122.x (2004).
26 Prada, J. A. H. *et al.* Ancient evolutionary origin of diversified variable regions demonstrated by crystal structures of an immune-type receptor in amphioxus. *Nat Immunol* **7**, 875-882, doi:http://www.nature.com/ni/journal/v7/n8/suppinfo/ni1359_S1.html (2006).
27 Arendt, D. *et al.* The origin and evolution of cell types. *Nat Rev Genet* **17**, 744-757, doi:10.1038/nrg.2016.127 (2016).
28 Simionato, E. *et al.* Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. *BMC Evolutionary Biology* **7**, 33, doi:10.1186/1471-2148-7-33 (2007).
29 Massari, M. E. & Murre, C. Helix-Loop-Helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol* **20**, doi:10.1128/mcb.20.2.429-440.2000 (2000).
30 Shain, D. H., Neuman, T. & Zuber, M. X. Embryonic Expression and Evolution of Duplicated E-Protein Genes in Xenopus laevis: Parallels With Ancestral E-Protein Genes. *Genetics* **146**, 345-353 (1997).
31 Diez-Roux, G. *et al.* A High-Resolution Anatomical Atlas of the Transcriptome in the Mouse Embryo. *PLOS Biology* **9**, e1000582, doi:10.1371/journal.pbio.1000582 (2011).
32 Semerad, C. L., Mercer, E. M., Inlay, M. A., Weissman, I. L. & Murre, C. E2A proteins maintain the hematopoietic stem cell pool and promote the maturation of myelolymphoid and myeloerythroid progenitors. *Proceedings of the National Academy of Sciences* **106**, 1930-1935, doi:10.1073/pnas.0808866106 (2009).
33 Welinder, E. *et al.* The transcription factors E2A and HEB act in concert to induce the expression of FOXO1 in the common lymphoid progenitor. *Proceedings of the National Academy of Sciences* **108**, 17402-17407, doi:10.1073/pnas.1111766108 (2011).
34 Barndt, R. J., Dai, M. & Zhuang, Y. Functions of E2A-HEB Heterodimers in T-Cell Development Revealed by a Dominant Negative Mutation of HEB. *Molecular and Cellular Biology* **20**, 6677-6685, doi:10.1128/mcb.20.18.6677-6685.2000 (2000).
35 Beck, K., Peak, M. M., Ota, T., Nemazee, D. & Murre, C. Distinct roles for E12 and E47 in B cell specification and the sequential rearrangement of immunoglobulin light chain loci. *The Journal of Experimental Medicine* **206**, 2271-2284, doi:10.1084/jem.20090756 (2009).
36 Dias, S., Månsson, R., Gurbuxani, S., Sigvardsson, M. & Kee, B. L. E2A Proteins Promote Development of Lymphoid-Primed Multipotent Progenitors. *Immunity* **29**, 217-227, doi:https://doi.org/10.1016/j.immuni.2008.05.015 (2008).

37 Ikawa, T., Kawamoto, H., Goldrath, A. W. & Murre, C. E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. *The Journal of Experimental Medicine* **203**, 1329-1342, doi:10.1084/jem.20060268 (2006).
38 Kwon, K. *et al.* Instructive Role of the Transcription Factor E2A in Early B Lymphopoiesis and Germinal Center B Cell Development. *Immunity* **28**, 751-762, doi:https://doi.org/10.1016/j.immuni.2008.04.014 (2008).
39 Takeuchi, A. *et al.* E2A and HEB Activate the Pre-TCRα Promoter During Immature T Cell Development. *The Journal of Immunology* **167**, 2157-2163, doi:10.4049/jimmunol.167.4.2157 (2001).
40 Wöhner, M. *et al.* Molecular functions of the transcription factors E2A and E2-2 in controlling germinal center B cell and plasma cell development. *The Journal of Experimental Medicine* **213**, 1201-1221, doi:10.1084/jem.20152002 (2016).
41 Wojciechowski, J., Lai, A., Kondo, M. & Zhuang, Y. E2A and HEB Are Required to Block Thymocyte Proliferation Prior to Pre-TCR Expression. *The Journal of Immunology* **178**, 5717-5726, doi:10.4049/jimmunol.178.9.5717 (2007).
42 Xu, W. *et al.* E2A transcription factors limit expression of Gata3 to facilitate T lymphocyte lineage commitment. *Blood* **121**, 1534-1542, doi:10.1182/blood-2012-08-449447 (2013).
43 Zhuang, Y., Cheng, P. & Weintraub, H. B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A, E2-2, and HEB. *Molecular and Cellular Biology* **16**, 2898-2905, doi:10.1128/mcb.16.6.2898 (1996).
44 Cisse, B. *et al.* Transcription Factor E2-2 Is an Essential and Specific Regulator of Plasmacytoid Dendritic Cell Development. *Cell* **135**, 37-48, doi:https://doi.org/10.1016/j.cell.2008.09.016 (2008).
45 Quong, M. W. *et al.* Receptor editing and marginal zone B cell development are regulated by the helix-loop-helix protein, E2A. *The Journal of experimental medicine* **199**, 1101-1112, doi:10.1084/jem.20031180 (2004).
46 Wikstrom, I., Forssell, J., Goncalves, M., Colucci, F. & Holmberg, D. E2-2 regulates the expansion of pro-B cells and follicular versus marginal zone decisions. *J Immunol* **177**, 6723-6729 (2006).
47 Han, Y. *et al.* The primitive immune system of amphioxus provides insights into the ancestral structure of the vertebrate immune system. *Developmental & Comparative Immunology* **34**, 791-796, doi:https://doi.org/10.1016/j.dci.2010.03.009 (2010).
48 Kasahara, M., Suzuki, T. & Pasquier, L. D. On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. *Trends in Immunology* **25**, 105-111, doi:https://doi.org/10.1016/j.it.2003.11.005 (2004).
49 Meister, M. Blood cells of Drosophila: cell lineages and role in host defence. *Current Opinion in Immunology* **16**, 10-15, doi:https://doi.org/10.1016/j.coi.2003.11.002 (2004).
50 Pancer, Z. Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. *Proceedings of the National Academy of Sciences* **97**, 13156-13161, doi:10.1073/pnas.230096397 (2000).
51 Zhang, B., Lin, Y.-Y., Dai, M. & Zhuang, Y. Id3 and Id2 Act as a Dual Safety Mechanism in Regulating the Development and Population Size of Innate-like γδ T Cells. *The Journal of Immunology* **192**, 1055-1063, doi:10.4049/jimmunol.1302694 (2014).
52 Schmitz, R. *et al.* Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* **490**, 116, doi:10.1038/nature11378

https://http://www.nature.com/articles/nature11378 - supplementary-information (2012).
53 Schrankel, C. S., Solek, C. M., Buckley, K. M., Anderson, M. K. & Rast, J. P. A conserved alternative form of the purple sea urchin HEB/E2-2/E2A transcription factor mediates a switch in E-protein regulatory state in differentiating immune cells. *Developmental Biology* **416**, 149-161, doi:https://doi.org/10.1016/j.ydbio.2016.05.034 (2016).
54 Hikima, J.-i. *et al.* Evolution of Transcriptional Control of the IgH Locus: Characterization, Expression, and Function of TF12/HEB Homologs of the Catfish. *The Journal of Immunology* **173**, 5476-5484, doi:10.4049/jimmunol.173.9.5476 (2004).
55 Ikawa, T., Kawamoto, H., Wright, L. Y. T. & Murre, C. Long-Term Cultured E2A-Deficient Hematopoietic Progenitor Cells Are Pluripotent. *Immunity* **20**, 349-360, doi:https://doi.org/10.1016/S1074-7613(04)00049-4 (2004).
56 King, A. M., Keating, P., Prabhu, A., Blomberg, B. B. & Riley, R. L. NK cells in the CD19− B220+ bone marrow fraction are increased in senescence and reduce E2A and surrogate light chain proteins in B cell precursors. *Mechanisms of Ageing and Development* **130**, 384-392, doi:https://doi.org/10.1016/j.mad.2009.03.002 (2009).
57 de Boer, J. *et al.* Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *European journal of immunology* **33**, 314-325, doi:10.1002/immu.200310005 (2003).
58 Bergqvist, I. *et al.* The basic helix-loop-helix transcription factor E2–2 is involved in T lymphocyte development. *European journal of immunology* **30**, 2857-2863, doi:doi:10.1002/1521-4141(200010)30:10<2857::AID-IMMU2857>3.0.CO;2-G (2000).
59 Pan, L., Hanrahan, J., Li, J., Hale, L. P. & Zhuang, Y. An analysis of T cell intrinsic roles of E2A by conditional gene disruption in the thymus. *J Immunol* **168**, 3923-3932 (2002).
60 Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Research* **46**, D754-D761, doi:10.1093/nar/gkx1098 (2018).
61 Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33**, 1870-1874, doi:10.1093/molbev/msw054 (2016).
62 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
63 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
64 Hayashi, S. & McMahon, A. P. Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse. *Dev Biol* **244**, 305-318, doi:10.1006/dbio.2002.0597 (2002).
65 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357, doi:10.1038/nmeth.1923
https://http://www.nature.com/articles/nmeth.1923 - supplementary-information (2012).

Acknowledgements

We thank: Prof. Joakim Dillner and colleagues (Karolinska Institutet) for access to the NextSeq 500 system; Associate Prof. Anders S. Nilsson (Stockholm University) for critical input on the phylogenetic analysis; Karolinska High Throughput Center (KHTC, Karolinska Institutet) for assistance with high-throughput sequencing; the preclinical laboratory at Karolinska University Hospital for access to their facilities; and UPPMAX Next Generation Sequencing Cluster & Storage (UPPNEX) for computational resources. This work was funded by the Swedish Cancer Foundation (Cancerfonden), the Swedish Research Council (VR), the Knut and Alice Wallenberg Foundation (KAW), the Swedish Foundation for Strategic Research (SSF), the European Commission (FP7 PIRG08-GA-2010-276906) and a generous donation by Björn and Lena Ulvaeus.

Authorship Contributions and conflict of interest statement

TB, CG and RM planned the study; TB, SK, XL, AK, MH, CG, KS and RM performed the experiments; TB, LPP, SK, XL, MH, ADP, CG, KS and RM analyzed the data; CS and DH provided critical insights; CS and RM supervised the research; TB, LPP, CG and RM wrote the manuscript; and all authors reviewed the manuscript before submission.

Competing interests

Authors have no conflict of interest to disclose.

Materials & Correspondence

Correspondence and request for material should be addressed to RM (robert.mansson@ki.se).

Figure legends

Figure 1. E2-2 and HEB are evolutionarily related and dispensable for ancestral lineages. (A) Schematic overview of evolutionary relationships and classification of animals included in the E-protein phylogenetic analysis. (B) Maximum Likelihood phylogenetic tree of the evolutionary distances between E-protein cDNA sequences in *Bilateria*. *Hydra vulgaris* was used as an out-group for the analysis. Colors indicate the E-protein families in *Gnathostomata*. Grey indicates branches with low support in the bootstrapping analysis (<70% of trees generated maintain the branches). For sequences used in the analysis, see Table S1. (C) Matrix displaying the pairwise p-distances between analyzed cDNA sequences. Absolute number of total cells (D), hematopoietic stem cells (HSCs) (E) and erythro-myeloid progenitors (F) in bone marrow. Absolute number of total cells (G), innate immune cells (H) (granulocytes, Gr; eosinophiles, Eos; macrophages, MF; and conventional dendritic cells, cDC) and plasmacytoid dendritic cells (pDC) (I) in spleen. For gating strategies for cell types in panel E-F and H see Fig. S3A-C. (J) Gating strategy for identification of pDCs. Color symbols utilized throughout the figure to indicate the genotype of analyzed mice are shown in the bottom right corner. Significance was calculated using the Mann-Whitney U test with *, ** and *** indicating p-values <0.05, <0.01 and <0.001 respectively.

Figure 2. E2-2 and HEB are needed for proper B-cell and CD4 T-cell generation. (A) Absolute number of T-cells, CD4 T-cells and CD8 T-cells in spleen. For T-cell gating strategy see Fig. S3C. Absolute number of B-cells (B) and B-cell subsets (C-D) (including: transitional 1-, T1; transitional 2-; marginal zone-, MZ; and follicular B-cells, FoB) in spleen. (E) Gating strategy for identification of B-cell subsets. Color symbols utilized throughout the figure to indicate the genotype of analyzed mice are shown below panels C-D. Significance was calculated using the Mann-Whitney U test with *, ** and *** indicating p-values <0.05, <0.01 and <0.001 respectively.

Figure 3. Maturation of CD4 and γδ T-cells is perturbed by the loss of E2-2 and HEB.

(A) Gating strategy for identification of T-cell developmental stages. Absolute numbers of: total cells (B); ETP/DN1 and DN2 (C); DN3E, DN3L, DN4, IS(8)P, DP TCRβ^- and DP TCRβ^+ (D); and γδ T-, CD4 T- and CD8 T-cells (E) in thymus. (F) Cell number ratios for each consecutive stage of development as a fraction of the prior stage. Each dot represents the ratio for one individual animal. Color symbols utilized throughout the figure to indicate the genotype of analyzed mice are shown at the bottom of the figure. Significance was calculated using the Mann-Whitney U test with *, ** and *** indicating p-values <0.05, <0.01 and <0.001 respectively.

Figure 4. The combined activity of E2-2 and HEB are critical for the generation of lymphoid progenitors and B-lineage cells. (A) Gating strategy for identification of stem- and lymphoid progenitors. (B-C) Absolute numbers of stem- and lymphoid progenitors indicated cell types in BM. Absolute numbers of total B-cells (D) and each B-cell developmental stage (E) in BM. (F) Gating strategy for identification of total B-cells and B-cell developmental stages. (G) Cell number ratios for each consecutive stage of development as a fraction of the prior stage. Each dot represents the ratio in an individual animal. Color symbols utilized throughout the figure to indicate the genotype of analyzed mice are shown below panel G. Significance was calculated using the Mann-Whitney U test with *, ** and *** indicating p-values <0.05, <0.01 and <0.001 respectively.

Figure 5. Loss of E-protein activity disrupts the early lymphoid transcriptional program.

(A) Principal component analysis of RNAseq data. (B) Hierarchically clustered heatmap showing the (row normalized) expression of genes with significant (adjusted p-value <0.01 calculated by EdgeR) changes in LY6D^{-}CLPs from E2-2$^{f/f}$Vav^{Cre} (Q1), E2-2$^{f/f}$HEB$^{f/f}$Vav^{Cre} (Q2) or E2A$^{f/f}$Vav^{Cre} (Q3) as compared to control (ctrl). Significance of expression changes in each comparison (Q1-3) is indicated to the right of the heatmap. (C) Venn diagrams showing the overlap between the significant expression changes in the indicated genotypes. Relative expression change (mean expression is set to one for each gene) of up-regulated (D) and down-regulated genes (E) from Q1-Q3. * indicates a p-value <0.002 calculated using the Mann-Whitney U test.

Figure 6. E2-2, HEB and E2A have partly overlapping binding patterns. (A) Number and overlap of identified E2-2, HEB and E2A peaks in proB cells. (B) Genomic localization of identified peaks. (C) Genome browser tracks showing E-protein binding near central B-lineage genes. (D) Percentage of genes in the E-protein dependent early lymphoid program (up/down regulated in Fig. 5B) and all expressed genes in LY6D^{-}CLPs that have HEB and E2A peaks within the gene body ±20kb. (E) Genome browser tracks showing E-protein binding near genes that were regulated by E-proteins in LY6D^{-}CLPs.

Fig. 1

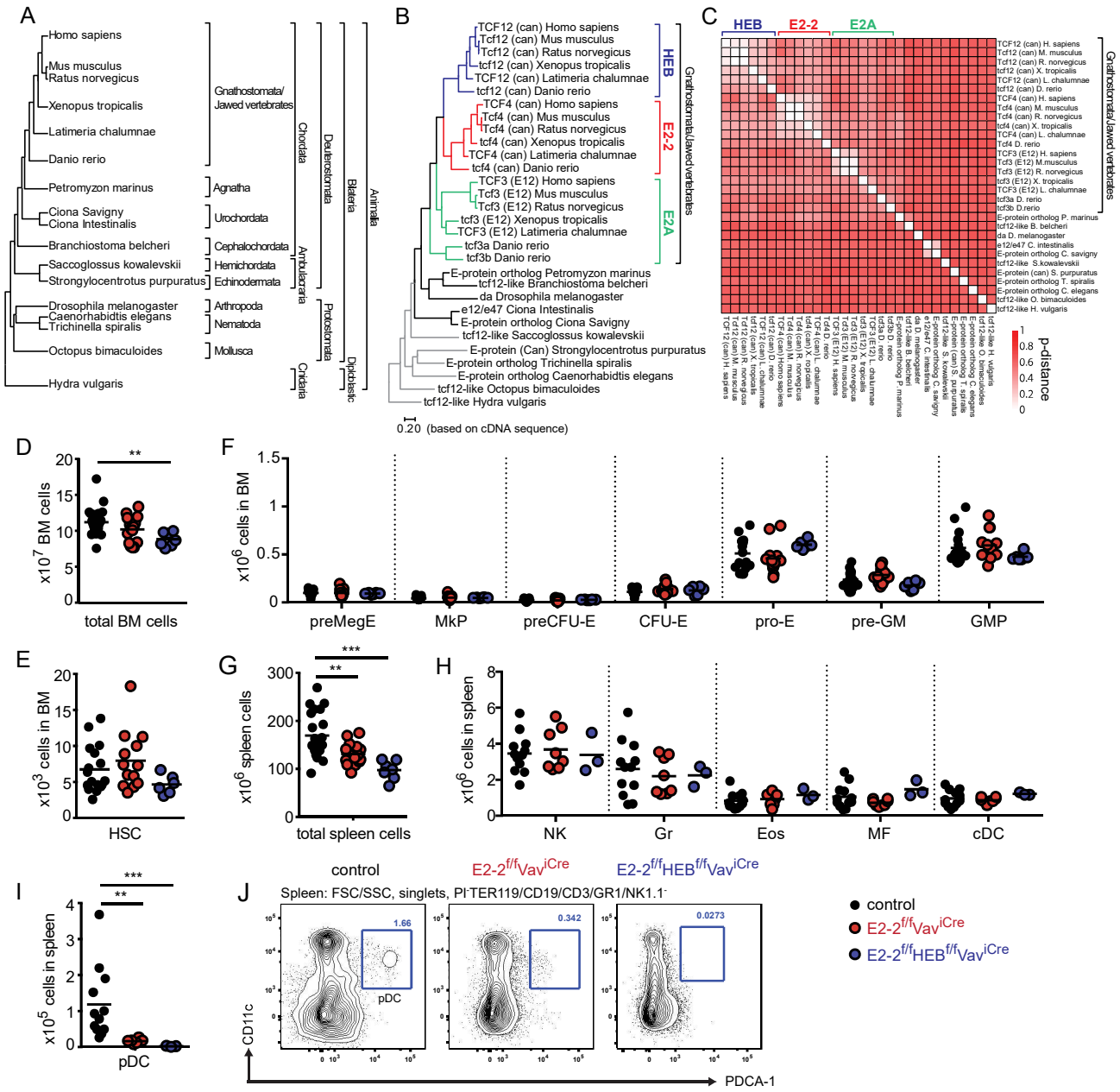

A
Homo sapiens
Mus musculus
Ratus norvegicus
Xenopus tropicalis
Latimeria chalumnae
Danio rerio
Petromyzon marinus
Ciona Savigny
Ciona Intestinalis
Branchiostoma belcheri
Saccoglossus kowalevskii
Strongylocentrotus purpuratus
Drosophila melanogaster
Caenorhabditis elegans
Trichinella spiralis
Octopus bimaculoides
Hydra vulgaris
Gnathostomata/ Jawed vertebrates
Agnatha
Urochordata
Cephalochordata
Hemichordata
Echinodermata
Arthropoda
Nematoda
Mollusca
Cnidaria
Chordata
Deuterostomata
Ambulacraria
Protostomata
Bilateria
Diploblastic
Animalia
B
TCF12 (can) Homo sapiens
Tcf12 (can) Mus musculus
Tcf12 (can) Ratus norvegicus
tcf12 (can) Xenopus tropicalis
TCF12 (can) Latimeria chalumnae
tcf12 (can) Danio rerio
TCF4 (can) Homo sapiens
Tcf4 (can) Mus musculus
Tcf4 (can) Ratus norvegicus
tcf4 (can) Xenopus tropicalis
TCF4 (can) Latimeria chalumnae
tcf4 (can) Danio rerio
TCF3 (E12) Homo sapiens
Tcf3 (E12) Mus musculus
Tcf3 (E12) Ratus norvegicus
tcf3 (E12) Xenopus tropicalis
TCF3 (E12) Latimeria chalumnae
tcf3a Danio rerio
tcf3b Danio rerio
E-protein ortholog Petromyzon marinus
tcf12-like Branchiostoma belcheri
da Drosophila melanogaster
e12/e47 Ciona Intestinalis
E-protein ortholog Ciona Savigny
tcf12-like Saccoglossus kowalevskii
E-protein (Can) Strongylocentrotus purpuratus
E-protein ortholog Trichinella spiralis
E-protein ortholog Caenorhabditis elegans
tcf12-like Octopus bimaculoides
tcf12-like Hydra vulgaris
HEB
E2-2
E2A
Gnathostomata/Jawed vertebrates
0.20 (based on cDNA sequence)
C
HEB
E2-2
E2A
Gnathostomata/Jawed vertebrates
p-distance
D
x10^7 BM cells
total BM cells
E
x10^3 cells in BM
HSC
F
x10^6 cells in BM
preMegE
MkP
preCFU-E
CFU-E
pro-E
pre-GM
GMP
G
x10^6 spleen cells
total spleen cells
H
x10^6 cells in spleen
NK
Gr
Eos
MF
cDC
I
x10^5 cells in spleen
pDC
J
control
E2-2f/fVavicCre
E2-2f/fHEBf/fVavicCre
Spleen: FSC/SSC, singlets, PI-TER119/CD19/CD3/GR1/NK1.1-
1.66
0.342
0.0273
pDC
CD11c
PDCA-1
control
E2-2f/fVavicCre
E2-2f/fHEBf/fVavicCre

Fig. 2

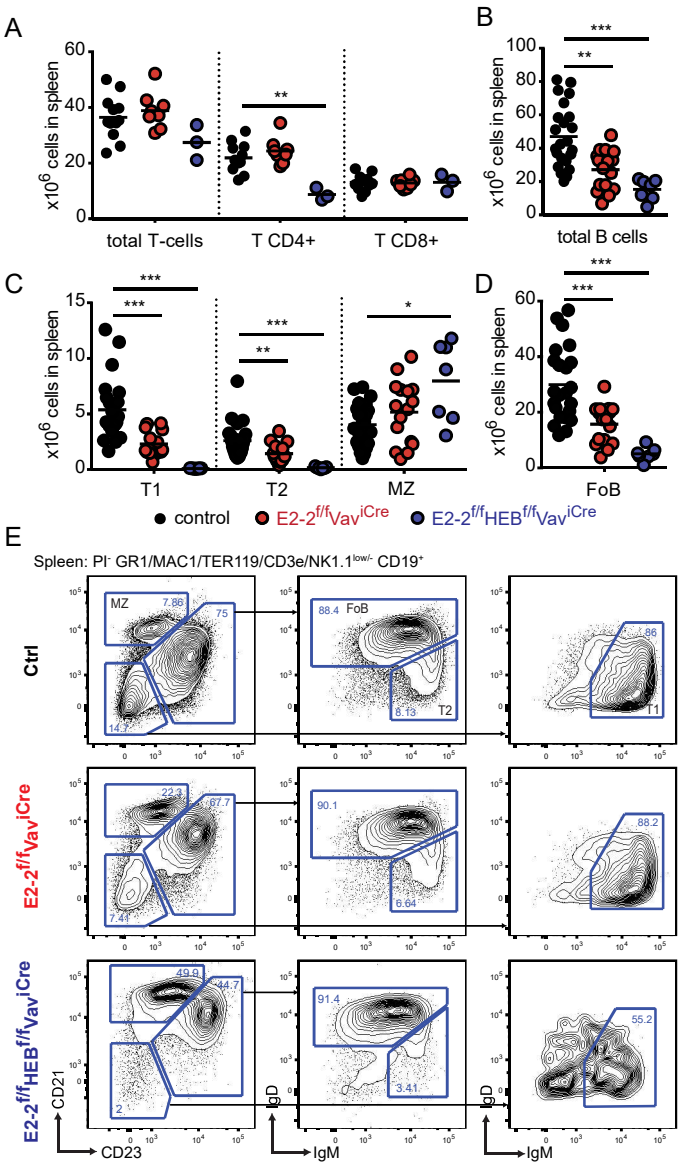

A
x10^6 cells in spleen
total T-cells
T CD4+
T CD8+
B
total B cells
C
T1
T2
MZ
D
FoB
control
E2-2f/fVaviCre
E2-2f/fHEBf/fVaviCre
E
Spleen: PI- GR1/MAC1/TER119/CD3e/NK1.1low/- CD19+
Ctrl
CD21
CD23
IgD
IgM

Fig. 3

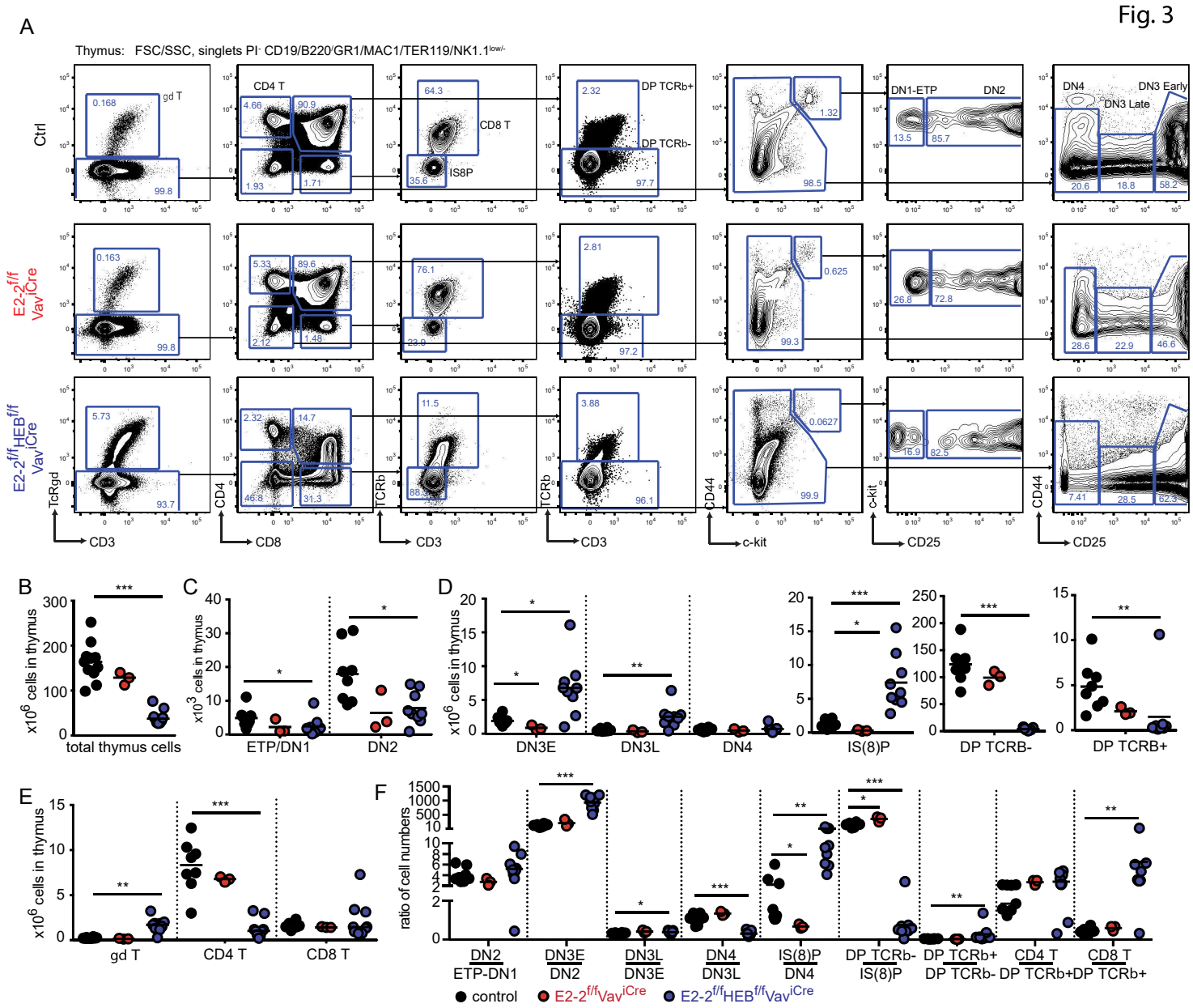

A
Thymus: FSC/SSC, singlets PI- CD19/B220-GR1/MAC1/TER119/NK1.1low/-
Ctrl
E2-2f/f VavicCre
E2-2f/fHEBf/f VavicCre
gd T
CD4 T
CD8 T
IS8P
DP TCRb+
DP TCRb-
DN1-ETP
DN2
DN4
DN3 Early
DN3 Late
TcRgd
CD4
TCRb
CD44
c-kit
CD3
CD8
CD25
B
x10^6 cells in thymus
total thymus cells
C
x10^3 cells in thymus
ETP/DN1
DN2
D
x10^6 cells in thymus
DN3E
DN3L
DN4
IS(8)P
DP TCRB-
DP TCRB+
E
x10^6 cells in thymus
gd T
CD4 T
CD8 T
F
ratio of cell numbers
DN2 ETP-DN1
DN3E DN2
DN3L DN3E
DN4 DN3L
IS(8)P DN4
DP TCRb- IS(8)P
DP TCRb+ DP TCRb-
CD4 T DP TCRb+
CD8 T DP TCRb+
control
E2-2f/fVavicCre
E2-2f/fHEBf/fVavicCre

Fig. 4

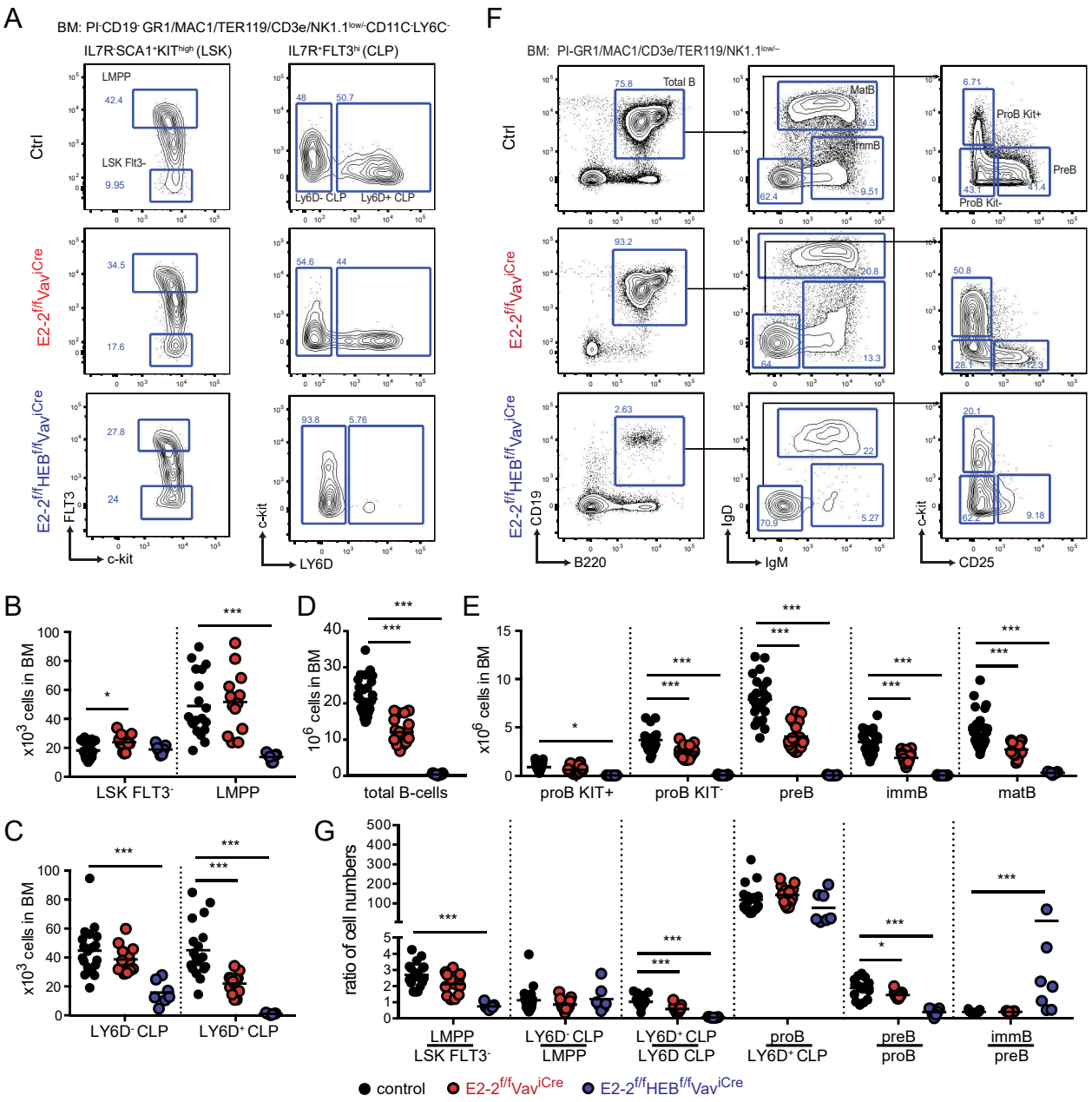

A
BM: PI-CD19-GR1/MAC1/TER119/CD3e/NK1.1low/-CD11C-LY6C-
IL7R-SCA1+KIThigh (LSK)
IL7R+FLT3hi (CLP)
Ctrl
E2-2f/fVavjCre
E2-2f/fHEBf/fVavjCre
LMPP
LSK Flt3-
Ly6D- CLP
Ly6D+ CLP
FLT3
c-kit
LY6D
F
BM: PI-GR1/MAC1/CD3e/TER119/NK1.1low/-
Total B
MatB
ImmB
ProB Kit+
ProB Kit-
PreB
CD19
B220
IgD
IgM
CD25
B
x10^3 cells in BM
LSK FLT3-
LMPP
D
x10^6 cells in BM
total B-cells
E
x10^6 cells in BM
proB KIT+
proB KIT-
preB
immB
matB
C
x10^3 cells in BM
LY6D- CLP
LY6D+ CLP
G
ratio of cell numbers
LMPP / LSK FLT3-
LY6D- CLP / LMPP
LY6D+ CLP / LY6D CLP
proB / LY6D+ CLP
preB / proB
immB / preB
control
E2-2f/fVavjCre
E2-2f/fHEBf/fVavjCre

Fig. 5

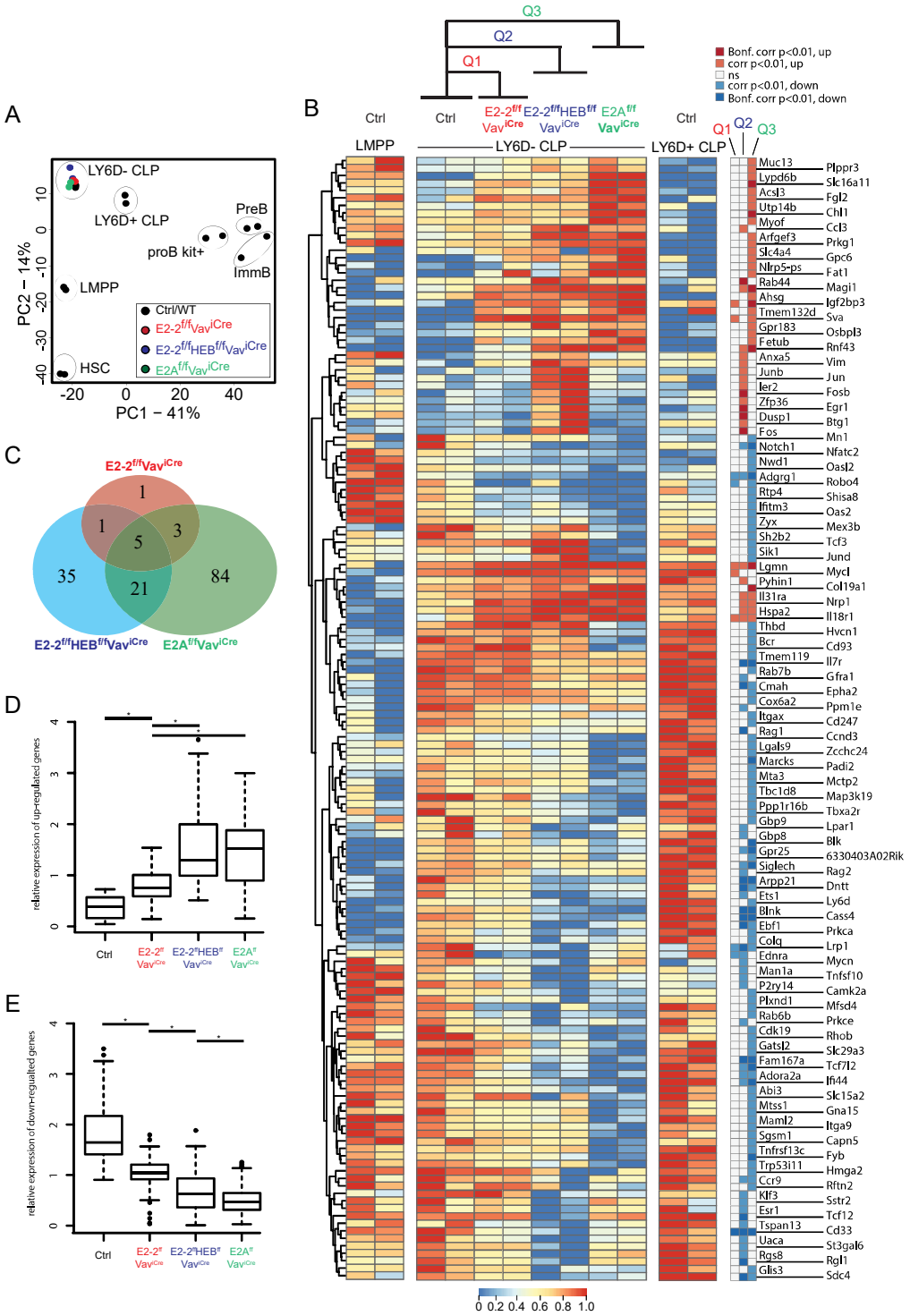
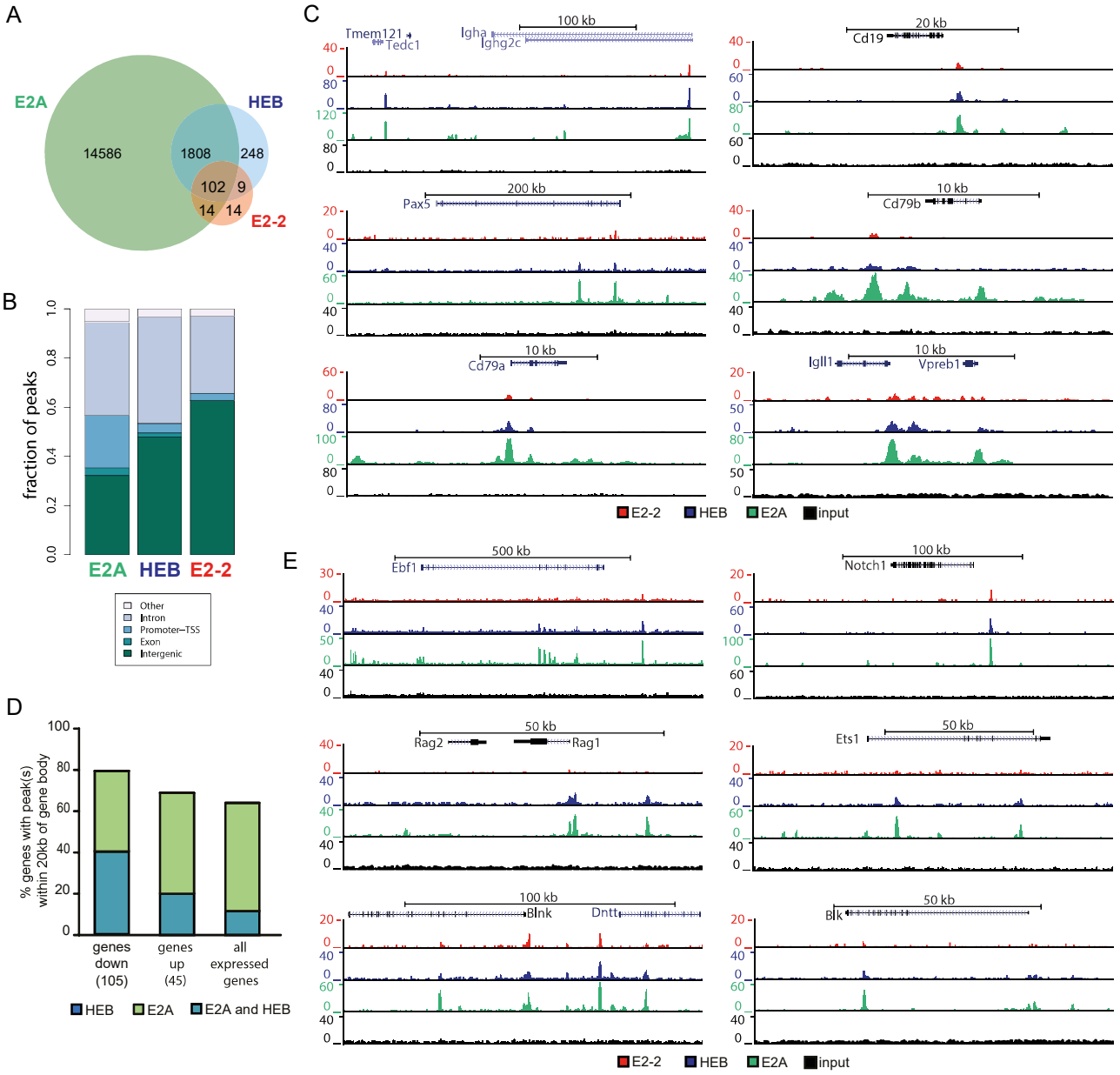

A
LY6D- CLP
LY6D+ CLP
PreB
proB kit+
ImmB
LMPP
HSC
Ctrl/WT
E2-2f/fVavicCre
E2-2f/fHEBf/fVavicCre
E2Af/fVavicCre
PC2 – 14%
PC1 – 41%
B
Q3
Q2
Q1
Bonf. corr p<0.01, up
corr p<0.01, up
ns
corr p<0.01, down
Bonf. corr p<0.01, down
Ctrl
LMPP
LY6D- CLP
LY6D+ CLP
Muc13
Lypd6b
Acsl3
Utp14b
Myof
Arfgef3
Slc4a4
Nlrp5-ps
Rab44
Ahsg
Tmem132d
Gpr183
Fetub
Anxa5
Junb
Ier2
Zfp36
Dusp1
Fos
Notch1
Nwd1
Adgrg1
Rtp4
Ifitm3
Zyx
Sh2b2
Sik1
Lgmn
Pyhin1
Il31ra
Hspa2
Thbd
Bcr
Tmem119
Rab7b
Cmah
Cox6a2
Itgax
Rag1
Lgals9
Marcks
Mta3
Tbc1d8
Ppp1r16b
Gbp9
Gbp8
Gpr25
Siglech
Arpp21
Ets1
Blnk
Ebf1
Colq
Ednra
Man1a
P2ry14
Plxnd1
Rab6b
Cdk19
Gatsl2
Fam167a
Adora2a
Abi3
Mtss1
Maml2
Sgsm1
Tnfrsf13c
Trp53i11
Ccr9
Klf3
Esr1
Tspan13
Uaca
Rgs8
Glis3
Plppr3
Slc16a11
Fgl2
Chl1
Ccl3
Prkg1
Gpc6
Fat1
Magi1
Igf2bp3
Sva
Osbpl3
Rnf43
Vim
Jun
Fosb
Egr1
Btg1
Mn1
Nfatc2
Oasl2
Robo4
Shisa8
Oas2
Mex3b
Tcf3
Jund
Mycl
Col19a1
Nrp1
Il18r1
Hvcn1
Cd93
Il7r
Gfra1
Epha2
Ppm1e
Cd247
Ccnd3
Zcchc24
Padi2
Mctp2
Map3k19
Tbxa2r
Lpar1
Blk
6330403A02Rik
Rag2
Dntt
Ly6d
Cass4
Prkca
Lrp1
Mycn
Tnfsf10
Camk2a
Mfsd4
Prkce
Rhob
Slc29a3
Tcf7l2
Ifi44
Slc15a2
Gna15
Itga9
Capn5
Fyb
Hmga2
Rftn2
Sstr2
Tcf12
Cd33
St3gal6
Rgl1
Sdc4
0 0.2 0.4 0.6 0.8 1.0
C
E2-2f/fVavicCre
1
1
3
5
35
21
84
E2-2f/fHEBf/fVavicCre
E2Af/fVavicCre
D
relative expression of up-regulated genes
Ctrl
E2-2f/f VavicCre
E2-2fHEBf/f VavicCre
E2Af/f VavicCre
E
relative expression of down-regulated genes
Ctrl
E2-2f/f VavicCre
E2-2fHEBf/f VavicCre
E2Af/f VavicCre

Fig. 6


A
E2A
HEB
E2-2
14586
1808
248
102
9
14
14
B
fraction of peaks
E2A
HEB
E2-2
Other
Intron
Promoter–TSS
Exon
Intergenic
C
Tmem121
Tedc1
Igha
Ighg2c
100 kb
20 kb
Cd19
200 kb
Pax5
10 kb
Cd79b
Cd79a
Igll1
Vpreb1
E2-2
HEB
E2A
input
D
% genes with peak(s) within 20kb of gene body
genes down (105)
genes up (45)
all expressed genes
HEB
E2A
E2A and HEB
E
500 kb
Ebf1
Notch1
50 kb
Rag2
Rag1
Ets1
Blnk
Dntt
Blk
E2-2
HEB
E2A
input