Selective deployment of transcription factor paralogs with submaximal strength facilitates gene regulation in the immune system

Ludovica Bruno1, Vijendra Ramlall 2, Romain A. Studer 3,10, Stephan Sauer1, David Bradley3, Gopuraja Dharmalingam4, Thomas Carroll4, Mohamed Ghoneim5,6, Michaël Chopin7, Stephen L. Nutt 2,7,8, Sarah Elderkin9, David S. Rueda 5,6, Amanda G. Fisher1, Trevor Siggers 2, Pedro Beltrao3 and Matthias Merkenschlager 1,14

In multicellular organisms, duplicated genes can diverge through tissue-specific gene expression patterns, as exemplified by highly regulated expression of RUNX transcription factor paralogs with apparent functional redundancy. Here we asked what cell-type-specific biologies might be supported by the selective expression of RUNX paralogs during Langerhans cell and inducible regulatory T cell differentiation. We uncovered functional nonequivalence between RUNX paralogs. Selective expression of native paralogs allowed integration of transcription factor activity with extrinsic signals, while non-native paralogs enforced differentiation even in the absence of exogenous inducers. DNA binding affinity was controlled by divergent amino acids within the otherwise highly conserved RUNT domain and evolutionary reconstruction suggested convergence of RUNT domain residues toward submaximal strength. Hence, the selective expression of gene duplicates in specialized cell types can synergize with the acquisition of functional differences to enable appropriate gene expression, lineage choice and differentiation in the mammalian immune system.

The majority of mammalian genes, including those for transcription factors, belong to gene families that have evolved following duplications of ancestral genes, genome segments or entire genomes 1,2. Gene duplications have been suggested as major drivers in the evolution of biological complexity because they provide redundancy that may allow for the accumulation of mutations; however, the mechanisms that govern the fate of duplicated genes are incompletely understood 3. While most duplicates are eliminated 4 or decrease their expression to match the dosage of the ancestral gene, others diverge by asymmetric tissue expression 5, often without showing differences in biochemical function 6. The duplication, degeneration and complementation (DDC) model 7 tries to explain how duplicated genes escape nonfunctionalization without the acquisition of new functions and therefore without selective advantage. The model suggests that complementary degenerative mutations in gene regulatory elements can increase the probability of duplicate gene preservation. What remains unclear is to what extent such tissue-specific expression patterns facilitate the evolution of new functions. To query what cell-type-specific biologies may emerge from the selective expression of apparently redundant transcription factor paralogs, we focused on the RUNX gene family. RUNX paralogs show tissue-specific expression, but have the same consensus DNA sequence 8 and can compensate for one another when expressed experimentally, indicating functional redundancy that appears limited only by their largely reciprocal expression 9,10. By quantitative analysis, we uncovered functional and biochemical differences between paralogs that were mediated by paralog-specific amino acids.

RUNX paralogs were selectively expressed during inducible regulatory T (Treg) cell and Langerhans (LH) cell differentiation, and the enforced expression of non-native paralogs interfered with physiological regulation by driving cell fate decisions in the absence of appropriate environmental signals. This observation suggests that endogenously expressed paralogs are of submaximal strength and is reminiscent of the use of submaximal transcription factor DNA binding motifs in developmental enhancers 11,12. Replacement of low-affinity motifs with high-affinity motifs can perturb developmental gene expression 13,14. Hence, there is no question that high-affinity DNA binding sites are important for transcription regulation 15, transcription factor binding sites of submaximal strength also make important contributions to the spatiotemporal regulation of gene expression in a range of developmental systems 16,17. Our data show that similar principles operate at the level of transcription factor protein sequences.

Functional differences between RUNX paralogs in regulation of the expression of Foxp3 and other RUNX target genes in T cells were explained by a small number of divergent amino acids within the otherwise highly conserved RUNT domain. Evolutionary reconstruction suggested convergence of RUNT domain residues toward submaximal DNA binding and reduced functional potency.

1Lymphocyte Development Group, MRC London Institute for Medical Sciences, London, UK. 2Department of Biology, Boston University, Boston, MA, USA. 3European Molecular Biology Laboratory (EMBL), European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK. 4MRC London Institute for Medical Sciences, London, UK. 5Molecular Virology, Department of Medicine, Imperial College London, London, UK. 6Single Molecule Imaging Group, MRC London Institute of Medical Sciences, London, UK. 7The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia. 8Department of Medical Biology, The University of Melbourne, Parkville, Victoria, Australia. 9Nuclear Dynamics Programme, The Babraham Institute, Cambridge, UK. 10Present address: BenevolentAI, London, UK. 11e-mail: matthias.merkenschlager@lms.mrc.ac.uk
Our data illustrate how the selective expression of gene duplicates in specialized cell types can synergize with the acquisition of functional differences to support the integration of extrinsic signals with endogenous transcription factor activities, supporting appropriate gene expression, lineage choice and differentiation in the mammalian immune system.

**Results**

**Origin and conservation of RUNX paralogs.** Runx1, Runx2 and Runx3 emerged from an ancestral gene during successive genome duplications that occurred near the root of the vertebrate tree (Fig. 1a). RUNX paralogs share the highly conserved RUNT domain (Fig. 1b) that associates with core binding factor-β (CBFβ) for high-affinity DNA binding. RUNX paralog expression is tissue specific and essential for osteogenesis, neurogenesis and definitive hematopoiesis. The expression of RUNX1 and RUNX3 is highly choreographed and often mutually exclusive during hematopoiesis, with programmed changes at key developmental transitions. Each RUNX paralog shows distinctive association with specific human diseases as a hallmark of regulatory neo- or subfunctionalization.

**RUNX paralogs in Langerhans cell differentiation.** To challenge the perception of functional redundancy between RUNX paralogs, we tested their ability to substitute for each other in the differentiation of LH cells, a skin-homing subset of dendritic cells (DCs). RUNX3 was expressed in immature DCs (Supplementary Fig. 1a, left), and is required for the transforming growth factor-β (TGF-β)-driven differentiation of LH cells, which are positive for CD11c, major histocompatibility complex (MHC) class II, DEC205 and epithelial cell adhesion molecule (EpCAM). LH cell differentiation is abolished by deleting the upstream regulator of Runx3 and Spi1 (PU.1), and can be rescued by RUNX3 expression in Spi1-deficient bone marrow (BM) cells. To address the equivalence of RUNX1 and RUNX3 we transduced conditionally Spi1-deficient BM cells with FLAG-RUNX1-IRES-GFP, FLAG-RUNX3-IRES-GFP or IRES-GFP control vector and sorted them into GFP-low, GFP-intermediate and GFP-high (lo/int/hi) cells. RUNX expression in these cells was quantified by immunoblotting for the FLAG epitope tag (Fig. 2a).

Spi1 deficiency abolished LH cell differentiation as reported. Both RUNX1 and RUNX3 were able to increase the number of Spi1-deficient CD11c+ cells expressing MHC class II. Unexpectedly, at matched levels of expression, the noncanonical paralog RUNX1 rescued LH cell differentiation more efficiently than the native RUNX3, as judged by the expression of EpCAM and DEC205 on CD11c+MHC class II+ cells (Fig. 2b and see Supplementary Fig. 1b for cell numbers). TGF-β is critical for LH cell differentiation in the presence of the native paralog RUNX3 (refs. 34,36). Interestingly, expression of the non-native paralog RUNX1 was able to override the requirement for TGF-β in the generation of LH cells (Fig. 2c and Supplementary Fig. 1b).
As observed for Spi1-deficient cells, RUNX1 also showed greater potency than RUNX3 in wild-type cells, both in the presence (Supplementary Fig. 2d,e) and absence (Supplementary Fig. 2f,g) of TGF-β (constructs representing transcripts from the proximal and distal Runx1 promoter33 were equally effective; data not shown). Therefore, the non-native paralog RUNX1 supported LH cell differentiation more efficiently than the endogenously expressed paralog RUNX3, and was able to override the requirement for TGF-β in the generation of LH cells.

RUNX paralogs in T cell differentiation. RUNX is required for the expression of Foxp3 (refs. 37–39), which encodes the signature transcription factor of Treg cells40. Unlike LH cells, Treg cells and their precursors preferentially express RUNX1 (refs. 31,32 and Supplementary Fig. 2a). To test the potency of the non-native paralog RUNX3 in Foxp3 regulation, naive (CD4+CD25+CD62Lhi) CD4+ T cells from Rosa26 Er2Cre Runx1fl/fl mice were depleted of preexisting Treg cells and activated, and 4-hydroxytamoxifen (4-OHT) was added to induce Cre-mediated deletion of endogenous Runx1. Cells were transduced with RUNX1-IRES-GFP or RUNX3-IRES-GFP and FOXP3 protein expression was assessed at the single-cell level. Reconstitution of Runx1-deficient naive CD4+ T cells with RUNX1 restored the generation of FOXP3-expressing Treg cells in response to TGF-β and a combination of phosphatidylinositol-3-OH kinase (PI(3)K) inhibitor LY294002 and the mTOR inhibitor rapamycin45 in a dose-dependent manner (Fig. 3a). At an equivalent expression level (Fig. 3b) RUNX3 consistently induced a higher proportion of FOXP3+ cells than RUNX1 (Fig. 3a and Supplementary Fig. 2b). In Runx1-deficient wild-type CD4+ T cells, RUNX3, but not RUNX1, further increased the frequency of FOXP3-expressing cells (Supplementary Fig. 2c). RUNX3 also promoted Foxp3 induction in the absence of TGF-β or PI(3)K and mTOR inhibitors (Fig. 3b).

Differences in Foxp3 regulation map to the RUNT domain. To test whether functional differences between the RUNX paralogs are encoded in the highly conserved RUNT domain or the more diverged N and C termini, we used the previously reported ability of isolated RUNT domains to dominantly interfere with endogenous RUNX proteins37,42 (see Supplementary Fig. 3a for sequences). At equivalent levels of expression, the RUNT domain of RUNX3 (RUNT3) but not RUNX1 (RUNT1) potently antagonized Foxp3 induction in activated CD4+ T cells (Fig. 4). The RUNT domain therefore encodes key differences between RUNX1 and RUNX3 in the regulation of Foxp3. In contrast, functional differences between RUNX paralogs in LH cell differentiation mapped outside the RUNT domain (Supplementary Fig. 3b).

Mechanisms that mediate differences between RUNT domains. To address the mechanistic basis for the dominant-negative activity of isolated RUNT domains in this system, we examined their cellular localization. Immunofluorescence staining and confocal

---

**Fig. 2 | Selective deployment of RUNX paralogs enables signal-responsive Langerhans cell differentiation.** a, Expression of FLAG–RUNX by immature DCs transduced with FLAG–RUNX1–IRES–GFP or FLAG–RUNX3–IRES–GFP and sorted into cells with low, intermediate or high GFP expression. Levels of retrovirally encoded RUNX3 protein were comparable to those of endogenous RUNX3 protein in wild-type immature DCs (Supplementary Fig. 1a, right). Lamin was the loading control. Data are representative of three independent experiments. b, Spi1-deficient BM cells were cultured with granulocyte–macrophage colony-stimulating factor and TGF-β or PI(3)K and mTOR inhibitors (Fig. 3b).RUNX3 consistently induced a higher proportion of Langerhans cells after 72 h. Right: mean ± s.d. of three independent experiments. See Supplementary Fig. 1b for numbers of cells recovered. *P < 0.05, **P < 0.01, ***P < 0.001 by two-tailed Student’s t test of comparisons between RUNX1 and RUNX3 (black), RUNX1 and control vector (blue) and RUNX3 and control vector (red). c, Spi1-deficient immature DCs were cultured without TGF-β and analyzed as in b. Right: mean ± s.d. of three independent experiments. Statistics are the same as in b.
microscopy showed that FLAG-tagged RUNT1 and RUNT3 were both predominantly nuclear in the absence of heterologous nuclear localization sequences (Fig. 5a), indicating that the greater regulatory potency of RUNT3 was not explained by differential nuclear localization.

We used FLAG chromatin immunoprecipitation (ChIP) and real-time PCR to quantify chromatin binding of isolated RUNT1 and RUNT3 domains expressed in T cells under the same conditions as the Foxp3 induction experiments described above, to map chromatin binding of RUNT1 and RUNT3 by ChIP. RUNT3 interacted more strongly than RUNT1 with canonical RUNX binding sites at the Tcrb enhancer and the Foxp3 promoter (Fig. 5b, left). To explore whether RUNT1 or RUNT3 was able to compete with and displace endogenous full-length Runx in vivo, we used antibodies that recognize an epitope outside the RUNT domain in the C-terminal domain of RUNX1, RUNX2 and RUNX3. These antibodies therefore selectively bind full-length RUNX and not isolated RUNT domains. RUNT3 displaced endogenous RUNX protein more effectively than RUNT1 from binding sites at the Tcrb enhancer and Foxp3 promoter (Fig. 5b, right).

We applied purified RUNT1:CBFβ1 and RUNT3:CBFβ1 proteins to oligonucleotide libraries on universal protein-binding microarrays. This approach confirmed that RUNT1 and RUNT3 had the same DNA motif preferences (Fig. 5c). Titration electrophoretic mobility shift assays (EMSA) showed that RUNT3:CBFβ1 bound a consensus RUNX motif in the Foxp3 promoter with higher affinity than RUNT1:CBFβ1 (Fig. 5d and Supplementary Fig. 4). These experiments were done with excess CBFβ1, so that the limiting interaction was with DNA, not CBFβ1. The higher DNA-binding affinity observed for RUNT3 is consistent with the stronger binding of RUNT3 to bona fide Runx sites in vivo, the ability of RUNT3 to compete with endogenous full-length Runx for chromatin binding in vivo and the stronger dominant negative effect of RUNT3 on Foxp3 induction. We conclude that the RUNT domains of RUNX1 and RUNX3 share DNA-binding specificity, but differ in DNA binding affinity both in vitro and in vivo.

Amino acids that mediate functional differences between RUNT domains. There are nine amino acid differences between the RUNT domains of RUNX1 and RUNX3 (Fig. 6a). As these were the only differences between the RUNT1 and RUNT3 constructs used in our experiments (Supplementary Fig. 3a), they must account for the observed functional and biochemical differences between RUNT1 and RUNT3. To pinpoint the molecular basis for these differences, we replaced individual RUNT3 amino acids with residues from RUNT1 and tested the resulting chimeric RUNT domains for interference with Foxp3 induction. Replacement of RUNT3 V123 with alanine, the corresponding residue in RUNT1, reduced the ability of RUNT3 to interfere with Foxp3 induction (Fig. 6b). The V123A substitution was also relevant in the context of full-length RUNX3, as it rendered RUNX3 inefficient at driving Foxp3 induction (data not shown).

To address whether the V123A substitution is also key for regulating other RUNX target genes, we activated CD4+ T cells in the presence of the histone deacetylase inhibitor MS-275, which promotes expression of the RUNX target genes Gzmb, Prf1 and Tbx21.
RT–PCR showed that both RUNT1 and RUNT3 antagonized Gzmb, Pfl and Tbx21 mRNA induction, but RUNT3 was markedly more efficient than RUNT1. The V123A substitution reduced interference by RUNT3 to the level of RUNT1 (Fig. 6c). Similarly to the V123A substitution, replacement of RUNT3 V168 with isoleucine, the corresponding residue in RUNT1, reduced the ability of RUNT3 to interfere with Foxp3 induction (Supplementary Fig. 5). Titration EMSA showed that the V123A and V168I substitutions reduced the DNA binding affinity of RUNT3:CBFβ1 complexes (Supplementary Fig. 4).

In reciprocal experiments, we introduced the RUNT3-specific amino acid, V123, into the weaker paralog RUNT1. The A123V substitution converted RUNT1 into a potent inhibitor of Foxp3 binding to a canonical RUNX binding site in the Foxp3 promoter by EMSA, which equates to substantially reduced potency (Fig. 6c). Hence, transfer of a single RUNT domain amino acid residue was sufficient to strengthen the weaker paralog.

In addition to V123A and V168I, we also tested the impact of substituting RUNT3 residues A59 and T157 by the correspondingly alanine in RUNT1. In contrast to V123A and V168I, the A59P and T157P substitutions did not weaken, but instead strengthened the dominant-negative effect of RUNT3 on Foxp3 induction (Supplementary Fig. 6), and the A59P and T157P substitutions offset the impact of V123A and V168I when combined (Supplementary Fig. 6). Hence, while a subset of RUNT3-specific amino acids contribute to the greater strength of RUNT3, this strength does not appear to be maximized, and is partially offset by other RUNT3-specific amino acids (Fig. 7a).

**RUNT domain evolution.** To link the identification of functionally antagonist RUNT domain residues to the evolution of RUNX paralogs we performed an ancestral sequence reconstruction analysis on the basis of EnsemblCompara alignment with manual addition of sequences (Supplementary Fig. 7a–c). The most likely ancestral amino acid residues in positions 59 and 157 were valine (posterior probability = 0.99) and isoleucine (posterior probability = 0.74; valine posterior probability = 0.26, respectively). The ancestral amino acid residue in positions 59 and 157 was almost certainly proline (posterior probability = 0.99 for position 59 and 1.00 for position 157; Supplementary Fig. 7c).

Following the whole-genome duplication at the root of the vertebrate tree, V123 was initially retained in RUNX1. V123 is still found in RUNX1 of cartilaginous fish with manually added sequences (Supplementary Fig. 7d for posterior probabilities), but was subsequently substituted by alanine in the ancestor of bony vertebrates (see Supplementary Fig. 7d for posterior probabilities). The V123A substitution reduced the binding affinity of Runx:CBFβ1 for the consensus RUNX motif in the Foxp3 promoter by EMSA, which equates to substantially reduced potency.
in functional assays (blue in Fig. 7b,c and see Supplementary Fig. 6 for functional activity). Conversely, I168 was probably substituted by valine in the ancestor of RUNX2 and RUNX3, conferring a slight increase in functional potency (red in Fig. 7b,c). Ancestral P59 was preserved in RUNX1 and RUNX2 but was substituted by alanine at the branch leading to RUNX3, slightly reducing the regulatory potency of the RUNX3 RUNT domain (ancestral RUNX2 and RUNX3 versus ancestral vertebrate RUNX3 in Fig. 7c). P157 was substituted for Thr in the tetrapod branch of RUNX3 evolution, leading to a further reduction in the potency of RUNX3 (tetrapod RUNX3 in Fig. 7c). In contrast to RUNX3, RUNX1 retained the ancestral P59 and P157 residues, which partially compensate for the impact of the V123A substitution in RUNX1 (Fig. 7c). Hence, with respect to the RUNT domain residues examined, sequence divergence resulted in an apparent convergence of function in higher vertebrates, and the RUNT domains of RUNX1, RUNX2 and RUNX3 in higher vertebrates are now weaker than the putative ancestral RUNT domain (Fig. 7d).

Owing to the modest posterior probability for position 168 (posterior probability I168 = 0.74), we tested the most likely alternative for this position (V168, posterior probability = 0.26), in the presence of P59 (posterior probability = 0.99) and P157 (posterior probability = 1.00). The resulting alternative ancestral RUNT domain (with P59 and P157; data in Supplementary Fig. 6) combines key residues of RUNX1 (P59 and P157) and RUNX3 (V123 and V168) into a RUNT domain with slightly stronger regulatory activity than the putative ancestral domain containing I168 (Supplementary Fig. 8b). In this alternative scenario, all amino acid changes that occurred during the evolution of ancestral RUNT domain residues 59, 123, 157 and 168 weakened RUNT domain regulatory activity, consistent with RUNT domain evolution toward submaximal strength.

Discussion

Our analysis shows that RUNX paralogs differ not only in their pattern of expression, but also in functional and biochemical properties. Owing to these functional differences, the selective expression

Fig. 6 | Identification of residues that functionally distinguish paralogous RUNT domains. a, RUNT1- and RUNT3-specific amino acids; numbers refer to the position in mouse RUNX1. With the exception of I168, none of these residues contact DNA. b, Replacement of RUNT3 V123 by the RUNT1 A123 residue (V123A) weakens the dominant-negative activity of RUNT3. Data are shown as the mean ± s.d. of three independent experiments. **P < 0.01, ***P < 0.001 by two-tailed Student’s t test of comparisons between RUNT and RUNT3 V123A. c, Substitution at V123 affects regulation of the RUNX target genes Gzmb, Prf1 and Tbx21. Data are shown as the mean ± s.d. of three independent experiments. Results for all RUNT domain constructs were significantly different from the control vector (P < 0.001 by two-tailed Student’s t test). **P < 0.01, ***P < 0.001 by two-tailed Student’s t test. NS, not significant. d, RUNT1 A123V is a more potent antagonist of Foxp3 induction in CD4 T cells than RUNT1 at matched levels of expression (as judged by FLAG–RUNT immunoblotting of GFP-lo/int/hi). Data are shown as the mean ± s.d. of three independent experiments. ***P < 0.001 by two-tailed Student’s t test for comparison between RUNT1 and RUNT1 A123V.
of RUNX paralogs can direct gene expression and ultimately lineage choice. In both differentiation systems investigated, the expression of non-native paralogs interfered with the ability to integrate appropriate cellular signals. This demonstration of a role for weak transcription factor paralogs in gene expression complements earlier work showing that transcription factor binding sites of submaximal strength contribute to spatiotemporal regulation of gene expression. Just as such instances do not argue against a role for affinity DNA binding in other settings, our data do not call into question the general importance of strong transcription factor paralogs. For example, the alternative GATA paralog, GATA-3, is an inefficient substitute for the native GATA-1 in erythropoiesis. Similarly, selective expression of RUNX1, the paralog with weaker DNA-binding affinity, in developing T cells is followed by a switch to RUNX3 expression in committed T helper type 1 (TH1) CD4+ T cells and in the CD8+ T cell lineage, where the RUNX target genes Tbx21, Gzmb and Prf1 come under the regulatory control of the stronger RUNX paralog.

Our data point to at least two mechanisms that can contribute to RUNX paralog strength. In LH cell differentiation, RUNX1 was the
more potent paralog. This property seems to be encoded outside the conserved RUNT domain and the underlying mechanisms remain to be explored. The potency of RUNX3 in regulating Foxp3, Tbx21, Gzmb and Prf1 was encoded within the RUNT domain, which facilitated the identification of paralog-specific amino acids that specify functional and biochemical differences. The RUNX3-specific residues V123 and V168 increased the DNA-binding affinity of RUND-CBFβ1 complexes. These residues do not contain DNA or CBFβ but may modulate changes in RUND domain conformation, which are known to occur upon binding to CBFβ and DNA. The in vitro DNA binding affinity of the RUNT domains correlated with chromatin binding in vivo, and with functional potency in the regulation of Foxp3 and other RUNX target genes in T cells. These data support a role of DNA-binding affinity in the regulatory potency of RUNX paralogs, but do not exclude additional mechanisms.

The DDC model explains how duplicate genes can be maintained by degenerative mutations in regulatory regions in the absence of functional differences and without evolutionary selection. This model provides an explanation as to why a higher than expected fraction of gene duplications survive without having to invoke adaptive forces. However, once such cell type-specific patterns of expression are established it is possible that, either through drift or selection, gene duplicates may more readily acquire cell type-specific functions. Our data illustrate how a combination of regulatory and functional differences between paralogs can allow for appropriate cell fate choices in the immune system.

In summary, cell type-specific expression of gene duplicates synergizes with the acquisition of functional differences to support the integration of extrinsic signals with endogenous transcription factor activities, thereby supporting appropriate gene expression, lineage choice and differentiation in the mammalian immune system.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41590-019-0471-5.

Received: 15 May 2018; Accepted: 16 July 2019; Published online: 26 August 2019

References
1. Teichmann, S. & Babu, M. M. Gene regulatory network growth by duplication. Nat. Genet. 36, 492–496 (2004).
2. Innan, H. & Kondrashov, F. The evolution of gene duplications: Classifying and distinguishing between models. Nat. Rev. Genet. 11, 97–108 (2010).
3. Ohno, S. Evolution by Gene Duplication (Allen & Unwin; Springer-Verlag, 1970).
4. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155 (2000).
5. Lan, X. & Pritchard, J. K. Coregulation of tandem duplicate genes by complementary, functional differences and without evolutionary selection. This process may more readily acquire cell type-specific functions. Our data illustrate how a combination of regulatory and functional differences between paralogs can allow for appropriate cell fate choices in the immune system.

In summary, cell type-specific expression of gene duplicates synergizes with the acquisition of functional differences to support the integration of extrinsic signals with endogenous transcription factor activities, thereby supporting appropriate gene expression, lineage choice and differentiation in the mammalian immune system.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41590-019-0471-5.

Received: 15 May 2018; Accepted: 16 July 2019; Published online: 26 August 2019

References
1. Teichmann, S. & Babu, M. M. Gene regulatory network growth by duplication. Nat. Genet. 36, 492–496 (2004).
2. Innan, H. & Kondrashov, F. The evolution of gene duplications: Classifying and distinguishing between models. Nat. Rev. Genet. 11, 97–108 (2010).
3. Ohno, S. Evolution by Gene Duplication (Allen & Unwin; Springer-Verlag, 1970).
4. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155 (2000).
5. Lan, X. & Pritchard, J. K. Coregulation of tandem duplicate genes by complementary, functional differences and without evolutionary selection. This process may more readily acquire cell type-specific functions. Our data illustrate how a combination of regulatory and functional differences between paralogs can allow for appropriate cell fate choices in the immune system.

In summary, cell type-specific expression of gene duplicates synergizes with the acquisition of functional differences to support the integration of extrinsic signals with endogenous transcription factor activities, thereby supporting appropriate gene expression, lineage choice and differentiation in the mammalian immune system.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41590-019-0471-5.

Received: 15 May 2018; Accepted: 16 July 2019; Published online: 26 August 2019

References
1. Teichmann, S. & Babu, M. M. Gene regulatory network growth by duplication. Nat. Genet. 36, 492–496 (2004).
2. Innan, H. & Kondrashov, F. The evolution of gene duplications: Classifying and distinguishing between models. Nat. Rev. Genet. 11, 97–108 (2010).
3. Ohno, S. Evolution by Gene Duplication (Allen & Unwin; Springer-Verlag, 1970).
4. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155 (2000).
5. Lan, X. & Pritchard, J. K. Coregulation of tandem duplicate genes by complementary, functional differences and without evolutionary selection. This process may more readily acquire cell type-specific functions. Our data illustrate how a combination of regulatory and functional differences between paralogs can allow for appropriate cell fate choices in the immune system.

In summary, cell type-specific expression of gene duplicates synergizes with the acquisition of functional differences to support the integration of extrinsic signals with endogenous transcription factor activities, thereby supporting appropriate gene expression, lineage choice and differentiation in the mammalian immune system.
40. Josefowicz, S. Z., Lu, L. F. & Rudensky, A. Y. Regulatory T cells: mechanisms of differentiation and function. *Ann. Rev. Immunol.* 30, 531–564 (2012).

41. Taniuchi, I. et al. Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. *Cell* **111**, 621–633 (2002).

42. Telfer, J. C., Hedblom, E. E., Anderson, M. K., Laurent, M. N. & Rothenberg, E. V. Localization of the domains in Runx transcription factors required for the repression of CD4 in thymocytes. *J. Immunol.* **172**, 4359–4370 (2004).

43. Berger, M. F. et al. Compact, universal DNA microarrays to comprehensively determine transcription factor binding site specificities. *Nat. Biotechnol.* **24**, 1429–1435 (2006).

44. Boucheron, N. et al. CD4+ T cell lineage integrity is controlled by the histone deacetylases HDAC1 and HDAC2. *Nat. Immunol.* **15**, 439–448 (2014).

45. Aken, B. L. et al. Ensembl 2017. *Nucleic Acids Res.* **45**, 635–642 (2017).

46. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

47. Nah, G. S., Lim, Z. W., Tay, B. H., Osato, M. & Venkatesh, B. Runx family genes in a cartilaginous fish, the elephant shark (*Callorhinchus milii*). *PLoS ONE.* **9**, e93816 (2014).

48. Takahashi, S. et al. GATA factor transgenes under GATA-1 locus control rescue germline GATA-1 mutant deficiencies. *Blood* **96**, 910–916 (2000).

49. Cruz-Guilloty, F. et al. Runx3 and T-box proteins cooperate to establish the transcriptional program of effector CTLs. *J. Exp. Med.* **206**, 51–59 (2009).

50. Yan, J., Liu, Y., Lukasik, S. M., Speck, N. A. & Bushweller, J. H. CBF allosterically regulates the Runx1 Runt domain via a dynamic conformational equilibrium. *Nat. Struct. Mol. Biol.* **11**, 901–906 (2004).

**Acknowledgements**

We thank A. Warren (University of Cambridge), F. Kondrashov (Institute of Science and Technology Austria), D. Odom (CRUK Cambridge), T. Warnecke, P. Sarkies, S. Santos and B. Lenhard for discussion and advice; N. Speck (University of Pennsylvania) for conditional Runx1 mutants; J. Elliott, B. Patel and T. Adejumo for cell sorting; P. Leung for assistance; D. Djehdoul for help with immunofluorescence; and L. Game and her team for sequencing. This work was funded by the Medical Research Council UK (M.M. and A.G.F.), by Wellcome (grant 099276/Z/12/Z to M.M.) and the NIH (R01AI116829 to T.S.).

**Author contributions**

L.B. performed experiments, made figures and contributed to writing; V.R. performed experiments and made figures; R.A.S. analyzed data, made figures and contributed to writing; S.S. did experiments; D.B. analyzed data; G.D. analyzed data and made figures; T.C. analyzed data and made figures; M.G. performed experiments and made figures; M.C. provided materials and performed experiments; S.L.N. provided materials, conceptualized and supervised work; S.E. conceptualized and supervised work and performed experiments; D.S.R. provided materials and supervised work; T.S. conceptualized and supervised work, analyzed data and made figures; A.G.F. conceptualized work and contributed to writing; P.B. conceptualized and supervised work; and M.M. conceptualized and supervised work, analyzed data, made figures and contributed to writing.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41590-019-0471-5.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.M.

**Peer review information:** Laurie Dempsey was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

Mouse strains, cell sorting, culture, viral infections and immunofluorescence. Mouse work was performed according to the Animals (Scientific Procedures) Act under the authority of project license no. PPL 70/7556 issued by the Home Office, UK. BM cells from wild-type C57BL/6–129 or Igxcr-Cre; Spi16 mice (ref. 31) were cultured in granulocyte–macrophage colony-stimulating factor (20 ng ml−1) as described36. Cells were transduced with mouse stem cell virus (MSCV)-based vectors by spin infection and kept for 48 h with or without TGF-β (5 ng ml−1). Cells were stained for CD11c (N418), MHC class II (M1415.2), EpCam (15G8) and DEC205 (NLDC205, all from BioLegend) and analyzed on an ARIA III or LSRII flow cytometer.

Naïve CD4+CD25−CD62L+ or CD4+CD25+ T cells were sorted by flow cytometry from the lymph nodes of wild-type BALB/c or C57BL/6 mice or Runx1−/−mice33 crossed with Rosa26-ERCre mice34 on a mixed C57BL/6–129 background. Cells were cultured at 105 cells per 2 ml in 10% FCS (Invitrogen) overnight before removal of beads as described37. After 24 h, activated T cells were retrovirally transduced with MSCV-based vectors by spin infection (90 min, 2500 rpm, 37°C, no polycrylene). Cells were replated at 5×106 cells per well 48 h after infection. TGF-β (1 ng ml−1) or L5294002 (10 μg ml−1) and rapamycin (25 nM) were added as indicated. Cells were fixed and permeabilized (ecto) and activated and stained for CD4 (Invitrogen) and Foxp3 (eBioScience) as described. Histone deacetylase inhibitor MS-275 (2 μM, Sigma) was added for 48 h where indicated and GFP+ cells were sorted for RNA extraction.

RNA was isolated by using RNA-Bee and reverse transcribed, and results from real-time qPCR with primers for Tcrb (Invitrogen, Life Technologies) were normalized to the geometric mean of Hprt and Ubc. Full-length Runx1 and Runx3 cDNAs and the Runx1 and Runx3 sequences encoding the RUNT domain were obtained by RT–PCR and mutagenesis was performed by splicing by overlap extension PCR38.

Immunoblotting was done as previously described39. Anti-RUNX1 and anti-RUNX3 (A488; cat. no. 23090 and 135248) or anti-FLAG (Sigma, cat. no. F3165) was used to detect endogenous and FLAG-tagged RUNX1 and RUNX3 proteins and anti-lamin was used as loading control (Santa Cruz Biotechnology).

For immunofluorescence, cells were fixed with 2% paraformaldehyde for 10 min, permeabilized with 0.4% Triton X-100 for 10 min, blocked with 10% normal goat serum for 30 min and incubated with anti-FLAG (Sigma, cat. no. F1804) and anti-lamin (C20, cat. no. sc6216, Santa Cruz Biotechnology) diluted 1:500 in 2.5% BSA in PBS for 1 h at 4°C. After two washes in 2.5% BSA in PBS, cells were incubated with 1:200 Alexa Fluor 647 donkey anti-mouse (Life Technologies Cat. no. A31371) and 1:400 Alexa Fluor 568 rabbit anti-goat (Life Technologies Cat. no. A11079) for 1 h at 20°C. Images were taken with an Olympus IX70 microscope using Micromanager software.

Chromatin immunoprecipitation. Cells were resuspended at 2×106 cells per milliliter in 0.3 M DSP (Pierce/Thermo, cat. no. 22855) and incubated for 45 min on a roller. Paraformaldehyde (16% solution, Pierce) was added for the last 10 min to a final concentration of 1%. The reaction was quenched by adding 1 M Tris (pH 7.5) to a final concentration of 50 mM and kept rotating for 3 min before washing twice. The pellet was resuspended in 600 μM RIPA buffer containing protease inhibitors before sonication. Chromatin was sonicated to a fragment size of 250–1000 bp using a Bioruptor (Diagenode, cat. no. B-123-336) or 25 μg of sonicated protein G Dynabeads (Invitrogen) and 5 μg of antibody overnight at 4°C. Immunocomplexes were washed twice in low-salt RIPA buffer (150 mM NaCl, 1 mM Tris–HCl (pH 7.5), 1 mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate and protease inhibitors), once in high-salt RIPA (same buffer but with 500 mM NaCl), once in LiCl RIPA (same buffer but with 250 mM LiCl) and once in LiCl RIPA (pH 8.0) and eluted at 65°C overnight in 0.1 M NaHCO3, 1% SDS, 50 μg ml−1 proteinase K and 5 μg ml−1 RNase A. DNA was analyzed by SYBR Green (Qiagen) real-time PCR using primers to Tcrb promoter.

Electrophoretic mobility shift assays for dissociation constant determination. Double-stranded 30-nt DNA (5′-TGTATGAAAGATGTTGTCATCGGAC CGCGTA) to which the RUNX binding site is underlined) was generated by melting complementary oligonucleotides at 95°C for 3 min and cooling to 4°C at a rate of −0.1°C s−1. In a 15-μl reaction, 2 pmol of double-stranded DNA was labeled with 32P using T4 polynucleotide kinase (NEB, cat. no. M0201) with 20 μCi of [γ-32P]ATP (PerkinElmer, cat. no. BL502Z) at 37°C for 2 h and purified by QiAquick Nucleotide Removal kit (Qiagen). Binding reactions containing purified protein samples and 1 fmol of probe were incubated at 20°C in 25 mM HEPES (pH 7.5), 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.2 mg ml−1 BSA and 0.025% Triton X-100 for 30 min. Binding reactions were electrophoresed at 70 V on a 6% non-denaturing gel (6% acrylamide, 0.225% bisacrylamide, 45 μM Tris base, 44 mM boric acid and 1 mM EDTA) at 20°C. Cells were dried onto 3-mm Whatman paper at 80°C for 1 h under vacuum and exposed to BAS storage phosphate-scintillant (GE, cat. no. 28-9564-75) for ~12 h. Screens were imaged using a Typhoon Trio scanner, and peak integration was done using ImageJ on the band representing the complex of interest. The data were fitted to a standard saturation binding equation: F = [P]·([P] + Ks). F is the fraction of the DNA probe bound by protein, [P] is protein concentration and Ks is the equilibrium dissociation constant.

Evolutionary analysis. The precomputed alignment and tree for the Runx1, Runx2 and Runx3 gene family (query ENSG00000206333) were extracted from EnsembleCompara release 88 (ref. 39). The EnsembleCompara tree was built using the TREESEAT algorithm40. Sequences from the cartilaginous shark species Scyllorhinus canicula (NCBI Gene IDs: Runx1, ABL68113.1; Runx2, ABL68115.1; Runx3, ABL68117.1) and Callorhinus milii (NCBI Gene IDs: Runx1, AHW58137.1; Runx2, AHW58142.1; Runx3, AHW58145.1) were manually placed in the phylogenetic tree, on the basis of taxonomical topology. The alignment was optimized with MAPF (model L-INS-i)-y. The alignment and tree were cleaned of short and spurious sequences with TrimAl-2 and Newick Utilities41, down to 188 sequences. Ancestral sequence reconstruction was based on this dataset and the 3 RUNX sequences and paralogs in vertebrates and another set of basal organisms. The tree was rooted using the protostomian species (nematodes and fruit fly). Branch lengths were re-estimated with PhyML42 and statistical support for nodes was assessed by the aRT method using chi-squared values43. CodeML, from the PAML package was used to infer the most likely amino acid sequence at different nodes, based on their posterior probability (empirical Bayes method). The evolutionary model used was LG (aARatefle = 1). The gamma shape parameter was estimated with an alpha of 0.5 and four categories of sites (fix_alpha = 0, alpha = 0.5 and ncatG = 4). RUNX sequence alignments were visualized in Jalview44.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Source data are available online for Figs. 1–7 and Supplementary Figs. 2, 5, 6 and 8.

References

51. Caton, M. L., Smith-Raska, M. R. & Reizis, B. Notch–RBP-J signaling controls the homeostasis of CD8+ dendritic cells in the spleen. J. Exp. Med. 204, 1653–1664 (2007).
52. Däck, A. et al. PU.1 regulates the commitment of adult hematopoietic progenitors and restricts granulopoiesis. J. Exp. Med. 201, 1487–1502 (2005).
53. Gauer, G. N. et al. Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. Blood 106, 494–504 (2005).
54. Seibler, J. et al. Rapid generation of inducible mouse mutants. Nucleic Acids Res. 31, e12 (2003).
55. Horton, R. M. In vitro recombination and mutagenesis of DNA: SOEing transcription factors. Methods Mol. Biol. 15, 251–261 (1991).
56. Tenno, M. et al. ChIPβ deficiency preserves Langerhans cell precursors by lack of selective TGFβ1 receptor signalling. J. Exp. Med. 21, 2933–2946 (2017).
57. Lin, H.-W., Chang, Y.-Y., Wong, M. L., Lin, J. W. & Chang, T. J. Functional analysis of virion host shutoff protein of pseudorabies virus. Virology 324, 412–418 (2004).
58. Berger, M. E. & Bulyk, M. L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nat. Protoc. 4, 393–411 (2009).
59. Dudley, A. M., Aach, J., Steffen, M. A. & Church, G. M. Measuring absolute expression with microarrays with calibrated reference sample and an extended signal intensity range. Proc. Natl Acad. Sci. USA 99, 7554–7559 (2002).

60. Schreiber, F., Patricio, M., Muffato, M., Pignatelli, M. & Bateman, A. TreeFam v9: a new website, more species and orthology-on-the-fly. Nucleic Acids Res. 42, D922–D925 (2014).

61. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780 (2013).

62. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973 (2009).

63. Junier, T. & Zdobnov, E. M. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics 26, 1669–1670 (2010).

64. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321 (2010).

65. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst. Biol. 55, 539–552 (2006).

66. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191 (2009).
Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided

Only common tests should be described solely by name; describe more complex techniques in the Methods section.

- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficients) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

- For all hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  *Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on [statistics for biologists](https://www.nature.com) contains articles on many of the points above.

Software and code

Policy information about availability of computer code

| Data collection | BD FACSdiva 8.0 |
|-----------------|-----------------|
|                 | GenePix Pro 7.2 |
|                 | Micromanager software was used for image acquisition on Olympus IX10 |
|                 | EnsembleCompara Release 88 |

| Data analysis   | FlowJo v10 |
|-----------------|------------|
|                 | PhyML     |
|                 | PAML package |
|                 | Hiji      |
|                 | Prism-GraphPad Software version 8.1.2 |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. Github). See the Nature Research [guidelines for submitting code & software](https://www.nature.com) for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

No high throughput data were generated in this study.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-list.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size | No statistical test was used to determine sample size. Three or more biological replicates per group (as indicated in the figure legends) were used to allow statistical inference.
---|---
Data exclusions | No data were excluded by the analysis except for the standard quality control filtering during sequence analysis.
Replication | Replication attempts were successful and results are represented as mean +/- SDV as indicated.
Randomization | No randomized method was required. Mice used for primary BM and T cell cultures were chosen according to phenotype.
Blinding | Investigators were not blinded to group allocation. We considered that blinding was not relevant to this study because we used quantitative assays for all experiments and the computational framework was identical for all samples and replicates.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ■   | Antibodies            |
| ■   | Eukaryotic cell lines |
| ■   | Palaeontology         |
| ■   | Animals and other organisms |
| ■   | Human research participants |
| ■   | Clinical data         |

### Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ■   | ChIP-seq              |
| ■   | Flow cytometry        |
| ■   | MRI-based neuroimaging |

### Antibodies

**Antibodies used**

Bone marrow staining for Langerhans cells: MHC Class II-BV421 (MS/114.15.2, #107608), CD11c-biotin (N418, #117303), EpCam-PECy7 (G8.8, #118216), DEC-205-APC (NLDC145, #138206) all from Biolegend. Lymph-node staining for naïve T cell sorting: CD4-PE/Cy5 (RM4-5, #10051), CD62L-PerCP (MEL-14, #104405), CD25-PE (PC61, #10207), all from Biolegend, anti Foxp3 (FJK-16S, PE, #12-5773-80, ebioscience). Western Blots: anti-Flag (Sigma, #F1804), anti-Runx1 (ab23890) and anti-Runx3 (ab135248) from abcam, anti-Lamin (C20, #32216) Santa Cruz Biotechnology. ChIP: anti-Runx1-3 (ab92336). In protein binding microarray experiment: anti-HIS tag (Sigma H1029), AlexaFluor 488-conjugated goat anti mouse IgG (Life Tech A11011), A488-conjugated goat anti glutathione S-transferase (Life Tech A11131). For immunofluorescence imaging: AlexaFluor 568 rabbit anti-goat (Life Tech A11079) and AlexaFluor 647 donkey anti-mouse (Life Tech A31571).

**Validation**

All antibodies are commercially available and have been tested in mouse fore the experiments they were used.

### Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research.

**Laboratory animals**

B10/c, C57Bl/6J, Ltgaxcre Spifl/fl, Runx1lox/lox, Rosa26-Frt2Cre mice were used. Specific pathogen free male and female laboratory bred mice of the indicated genotype were used at 6-8 weeks of age.

**Wild animals**

No Wild animals were used in this study.

**Field-collected samples**

This study did not involve field-collected samples.
Flow Cytometry

Plots

- Confirm that:
  - The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
  - The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
  - All plots are contour plots with outliers or pseudocolor plots.
  - A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Bone marrow was flushed out from femur and tibia, cells were counted and cultured in GM-CSF +/- TGF-b. After retroviral transduction, cells were stained with MHC class II, CD11c, Dec-205 and EpCAM before being analyzed. Lymph node cells were isolated from inguinal, axillary, mandibular and mesenteric lymph nodes, stained with anti-CD4, CD25 and CD62L monoclonal antibodies, filtered and sorted. For Foxp3 nuclear staining, we used eBioscience kit (00-5523-00), following the manufacturer protocol.

Instrument

8D LSR II for and FACS Calibur for analysis, FACS Aria Fusion, Aria II and Aria III for sorting

Software

- 8D FACSDiva 8.0
- FlowJo v10
- Prism-GraphPad Software version 8.1.2

Cell population abundance

CD4+ CD25- lymph node cells were 23-27% of gated live, doublets excluded, parent population. Of these, the top 70% CD62L+ was sorted. Sorted cells were analyzed and purity was >99%.

Gating strategy

Cells were first gated on live cells (FSC/SSC) and gated to exclude doublets. To obtain naive T cells, samples were first gated on CD4+ CD25- cells and subsequently on CD62Lhi. Retro-virally transduced cells were sorted according to lower-third, middle-third and upper-third of GFP expression or for GFP+. Sorted cells were re-analysed to check for purity.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.