

Published in final edited form as:

*Nat Immunol.* 2019 October ; 20(10): 1372–1380. doi:10.1038/s41590-019-0471-5.

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

Ludovica Bruno<sup>1</sup>, Vijendra Ramlall<sup>2</sup>, Romain A. Studer<sup>3</sup>, Stephan Sauer<sup>1</sup>, David Bradley<sup>3</sup>, Gopuraja Dharmalingam<sup>4</sup>, Thomas Carroll<sup>4</sup>, Mohamed Ghoneim<sup>5,6</sup>, Michaël Chopin<sup>7</sup>, Stephen L. Nutt<sup>7</sup>, Sarah Elderkin<sup>8</sup>, David S. Rueda<sup>5,6</sup>, Amanda G. Fisher<sup>1</sup>, Trevor Siggers<sup>2</sup>, Pedro Beltrao<sup>3</sup>, Matthias Merkenschlager<sup>1,@</sup>

<sup>1</sup>Lymphocyte Development Group, MRC London Institute for Medical Sciences, London W12 0NN, UK

<sup>2</sup>Boston University, Boston, MA 02215, USA

<sup>3</sup>European Molecular Biology Laboratory (EMBL), European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK

<sup>4</sup>MRC London Institute for Medical Sciences, London W12 0NN, UK

<sup>5</sup>Molecular Virology, Department of Medicine, Imperial College London, London W12 0NN, UK

<sup>6</sup>Single Molecule Imaging Group, MRC London Institute of Medical Sciences, London W12 0NN, UK

<sup>7</sup>The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC 3052, Australia; Department of Medical Biology, The University of Melbourne, Parkville, VIC 3010, Australia

<sup>8</sup>Nuclear Dynamics Programme, The Babraham Institute, Cambridge, UK

### Abstract

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

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

@Address correspondence to [matthias.merkenschlager@lms.mrc.ac.uk](mailto:matthias.merkenschlager@lms.mrc.ac.uk). RAS's current address is BenevolentAI, London W1T 5HD, UK

**Author contributions.** LB did experiments, made figures and contributed to writing, VR did experiments and made figures, RAS analysed data, made figures and contributed to writing, SS did experiments, DB analysed data, GD analysed data and made figures, TC analysed data and made figures, MG did experiments and made figures, MC provided materials and did experiments, SLN provided materials, conceptualised and supervised work, SE conceptualised and supervised work and did experiments, DSR provided materials and supervised work, TS conceptualised and supervised work, analysed data and made figures, AGF conceptualised work and contributed to writing, PB conceptualised and supervised work, MM conceptualised and supervised work, analysed data, made figures, and contributed to writing.

### Competing interests statement

The authors declare that they have no competing interests

regulatory T cell differentiation. We uncovered functional non-equivalence 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 towards sub-maximal 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.

## Introduction

The majority of mammalian genes, including transcription factors, belong to gene families that have evolved following duplications of ancestral genes, genome segments, or entire genomes<sup>1,2</sup>. 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<sup>3</sup>. However, the mechanisms that govern the fate of duplicated genes are incompletely understood<sup>2</sup>. While most duplicates are eliminated<sup>4</sup> or decrease their expression to match the dosage of the ancestral gene<sup>5</sup>, others diverge by asymmetric tissue expression<sup>5</sup>, often without showing differences in biochemical function<sup>6</sup>. The duplication, degeneration and complementation (DDC) model<sup>7</sup> tries to explain how duplicated genes escape non-functionalization 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 share the same consensus DNA sequence<sup>8</sup> and can compensate for one another when expressed experimentally, indicating functional redundancy that appears limited only by their largely reciprocal expression<sup>9–11</sup>. 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 ( $T_{reg}$ ) 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<sup>12–21</sup>. Replacement of low- with high-affinity motifs can perturb developmental gene expression<sup>12–22</sup>. Hence, while there is no question that high-affinity DNA binding sites are important for transcriptional regulation<sup>23</sup>, transcription factor binding sites of submaximal strength also make important contributions to spatiotemporal gene expression in a range of developmental

systems<sup>12–22</sup>. Our data show that similar principles operate at the level of transcription factor protein sequences.

Functional differences between Runx paralogs in the regulation of *Foxp3* expression 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 towards submaximal DNA binding and reduced functional potency. 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 homology domain (RUNT domain, Fig. 1b) that associates with core binding factor beta (CBFB) for high-affinity DNA binding<sup>24–25</sup>. Runx paralog expression is tissue-specific and essential for osteogenesis<sup>26–27</sup>, neurogenesis<sup>28</sup>, and definitive hematopoiesis<sup>29–31</sup>. The expression of Runx1 and Runx3 is highly choreographed and often mutually exclusive during hematopoiesis, with programmed changes at key developmental transitions<sup>31–32</sup>. Each Runx paralog shows distinctive association with specific human diseases<sup>9,27,28,33</sup> as a hallmark of regulatory neo- or subfunctionalization<sup>17</sup>.

### 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<sup>34, 35</sup> (DCs). Runx3 is expressed in immature DCs (Supplementary Fig. 1a, left), and required for the transforming growth factor- $\beta$  (TGF- $\beta$ )-driven differentiation of CD11c<sup>+</sup>, MHC class II<sup>+</sup>, DEC205<sup>+</sup>, Epcam<sup>+</sup> LH cells<sup>34, 35</sup>. LH cell differentiation is abolished by deleting the upstream regulator of *Runx3*, *Spi1* (PU.1), and can be rescued by Runx3 expression in *Spi1*-deficient bone marrow (BM) cells<sup>35</sup>. 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 GFP-lo/int/hi cells. Runx expression in GFP-lo/int/hi cells was quantified by immunoblotting for the FLAG epitope tag (Fig. 2a). *Spi1*-deficiency abolished LH cell differentiation as reported<sup>34, 35</sup>. Both Runx1 and Runx3 were able to increase the number (Supplementary Fig. 1b) and the proportion of *Spi1*-deficient CD11c<sup>+</sup> cells expressing MHC-II (Supplementary Fig. 1c). Unexpectedly, at matched levels of expression, the non-canonical paralog Runx1 rescued LH cell differentiation more efficiently than the native Runx3 as judged by the expression of Epcam and DEC205 on CD11c<sup>+</sup> MHC-II<sup>+</sup> cells (Fig. 2b, see Supplementary Fig. 1b for cell numbers). TGF- $\beta$  is critical for LH cell differentiation in the presence of the native paralog Runx3 (ref. 34, 36). Interestingly, expression of the

non-native paralog Runx1 was able to override the requirement for TGF- $\beta$  in the generation of LH cells (Fig. 2c, Supplementary Fig. 1b).

As observed for *Sp1*-deficient cells, Runx1 showed greater potency than RUNX3 also in wild-type cells, both in the presence (Supplementary Fig. 2d, e) and in the absence of TGF- $\beta$  (Supplementary Fig. 2f, g, constructs representing transcripts from the proximal and distal *Runx1* promoter<sup>33</sup> 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- $\beta$  in the generation of LH cells.

### Runx paralogs in T cell differentiation

Runx is required for the expression of Foxp3 (ref. 37–39), the signature transcription factor of T<sub>reg</sub> cells (ref. 40). Unlike LH cells, T<sub>reg</sub> cells and their precursors preferentially express Runx1 (ref. 31, 41, Supplementary Fig. 2a). To test the potency of the non-native paralog Runx3 in Foxp3 regulation, naive (CD4<sup>+</sup> CD25<sup>-</sup> CD62Lhi) *Runx1*<sup>lox/lox</sup> ERT2Cre CD4<sup>+</sup> T cells were depleted of pre-existing T<sub>reg</sub> cells, activated, and 4-hydroxytamoxifen (4-OHT) was added to induce Cre-mediated deletion of endogenous *Runx1*. Cells were transduced with Runx1- or Runx3-IRES-GFP, and Foxp3 protein expression was assessed at the single-cell level. Reconstitution of *Runx1*-deficient naive CD4<sup>+</sup> T cells with Runx1 restored the generation of Foxp3-expressing T<sub>reg</sub> cells in response to TGF- $\beta$  and the combination of phosphatidylinositol-3-OH kinase (PI(3)K) inhibitor LY294002 and mTOR inhibitor rapamycin<sup>35</sup> in a dose-dependent manner (Fig. 3a). At equivalent expression (Fig. 3b) Runx3 consistently induced a higher proportion of Foxp3<sup>+</sup> cells than Runx1 (Fig. 3a, Supplementary Fig. 2b). In *Runx1*<sup>+/+</sup> CD4<sup>+</sup> 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- $\beta$  or PI(3)K and mTOR inhibitors (Fig. 3b).

These data show that Runx paralogs have distinct functional properties. Remarkably, in both LH cell and inducible T<sub>reg</sub> cell differentiation, the non-canonical Runx paralog was biologically more potent than the canonical Runx paralog, even in the absence of exogenous factors normally required to promote LH and T<sub>reg</sub> cell differentiation. Hence, the selective expression of transcription factor paralogs contributes to the regulation of cell fate choices.

### Differences in Foxp3 regulation map to the RUNT domain

To test whether functional differences between Runx paralog functions are encoded in the highly conserved RUNT domain or the more diverged N- or C-termini we used the previously reported ability of isolated RUNT domains to dominantly interfere with endogenous Runx proteins<sup>37, 42</sup>(see Supplementary Fig. 3a for sequences). At equivalent levels of expression, the RUNT domain of Runx3 (RUNT3) but not of Runx1 (RUNT1) potently antagonized Foxp3 induction in activated CD4<sup>+</sup> 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 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<sup>37</sup> at the *Tcrb* enhancer and the *Foxp3* promoter (Fig. 5b, left). To explore whether RUNT1 or RUNT3 were 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, 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:CBFB1 and RUNT3:CBFB1 proteins to oligonucleotide libraries on universal protein-binding microarrays<sup>43</sup>. This approach confirmed that both RUNT1 and RUNT3 shared the same DNA motif preferences (Fig. 5c). Titration EMSAs showed that RUNT3:CBFB1 bound a consensus Runx motif in the *Foxp3* promoter with higher affinity than did RUNT1:CBFB1 (Fig. 5d, Supplementary Fig. 4). These experiments were done with excess CBFB1, so that the limiting interaction was with DNA, not CBFB1. 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 9 amino acid differences between the RUNT domains of Runx1 and Runx3 (Fig. 6a). As these are 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 valine123 with alanine, the corresponding residue in RUNT1, reduced the ability of RUNT3 to interfere with *Foxp3* induction (Fig. 6b). The V123A

substitution was relevant also 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<sup>+</sup> T cells in the presence of the histone deacetylase inhibitor MS-275, which promotes the expression of the RUNX target genes *Gzmb*, *Prfl*, and *Tbx21* (ref. 44). RT-PCR showed that both RUNT1 and RUNT3 antagonized *Gzmb*, *Prfl*, 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). Similar to V123A, 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 EMSAs showed that the V123A and V168I substitutions reduced the DNA binding affinity of RUNT3:CBFB1 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* induction (Fig. 6d). 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 corresponding P59 and P157 residues from RUNT1. In contrast to V123A and V168I, A59P and T157P did not weaken, but instead strengthened the dominant-negative effect of RUNT3 on *Foxp3* induction (Supplementary Fig. 6), and A59P and T157P 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 antagonistic RUNT domain residues to the evolution of Runx paralogs we performed an ancestral sequence reconstruction analysis, based on EnsemblCompara alignment<sup>45, 46</sup> with manual addition of sequences (Supplementary Fig. 7a-c). The most likely ancestral amino acid residues in position 123 and 168 of the RUNT domain are Valine (posterior probability = 0.99) and Ile (probability = 0.74, Val = 0.26). The ancestral amino acid residue in positions 59 and 157 was almost certainly Proline (probability = 0.99 for P59 and 1.00 for P157, 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<sup>47</sup> (see Supplementary Fig. 7d for posterior probabilities), but was subsequently substituted by Ala in the ancestor of bony vertebrates (see Supplementary Fig. 7d for posterior probabilities). The V123A substitution reduced the binding affinity of Runx:CBFB1 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, see Supplementary Fig. 6 for functional activity). Conversely, I168 was likely 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 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 Threonine 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 ancestral P59 and P157, 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 higher vertebrate Runx1, Runx2 and Runx3 are now weaker than the putative ancestral RUNT domain (Fig. 7d)

Due 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 P = 0.99) and P157 (posterior probability P = 1.00). The resulting alternative ancestral RUNT domain (P59 P157, data in Supplementary Fig. 6) combines key residues of Runx1 (P59 P157) and Runx3 (V123, 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 towards submaximal strength.

## Discussion

Our analysis shows that Runx paralogs differ not only in their pattern of expression, but also in functional and biochemical properties. Due to these functional differences, the selective expression of Runx paralogs can direct gene expression, and ultimately lineage choice. Interestingly, in both differentiation systems investigated here, the expression of non-native paralogs interfered with the ability to integrate appropriate cellular signals.

The demonstration of a role for weak transcription paralogs in gene expression provided here complements earlier work showing that transcription factor binding sites of submaximal strength contribute to spatiotemporal gene expression<sup>12–22</sup>. Just as such instances do not argue against a role for affinity DNA binding in other settings<sup>23</sup>, our data do not 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<sup>48</sup>. Similarly, the selective expression of Runx1, the paralog with weaker DNA binding affinity, in developing T cells is followed by a switch to Runx3 in committed T<sub>H</sub>1 CD4<sup>+</sup> T cells and in the CD8<sup>+</sup> T cell lineage where the Runx target genes *Tbx21*, *Gzmb* and *Prf1* come under the regulatory control of the stronger Runx paralog<sup>49</sup>.

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 appears 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. Runx3-specific residues V123 and V168 increased the DNA binding affinity of RUNT:CBFB1 complexes. These residues do not contact DNA or CBFB (ref. 22, 23) but may modulate changes in RUNT domain conformation that are known to occur upon binding of CBFB and DNA (Ref. 50). In vitro DNA binding affinity of 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<sup>7</sup>. 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 synergises with the acquisition of functional differences to support the integration of extrinsic signals with endogenous transcription factor activities to support appropriate gene expression, lineage choice and differentiation in the mammalian immune system.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## 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 (UPenn) for conditional *Runx1* mutants, J. Elliott, B. Patel and T. Adejumo for cell sorting, P. Leung for assistance, D. Djeghloul for help with immunofluorescence, and L. Game and her team for sequencing. We thank the anonymous referees for their valuable input, in particular the suggestion to include cartilaginous fish in the analysis of RUNT domain evolution. This work was funded by the Medical Research Council UK and by Wellcome (099276/Z/12/Z to MM).

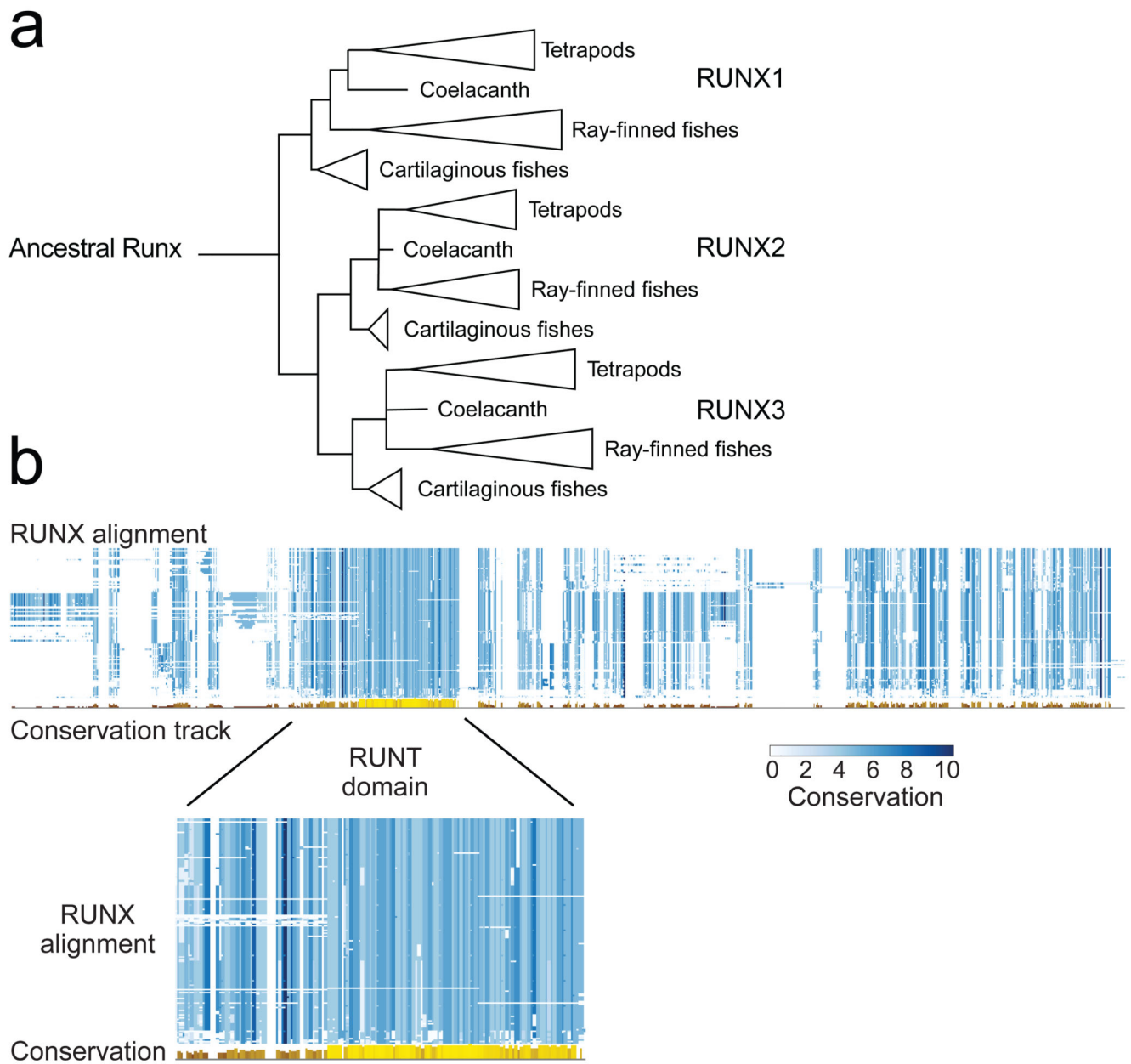
## References

1. Teichmann S, Babu MM. Gene regulatory network growth by duplication. *Nature Genetics*. 2004; 36:492–496. [PubMed: 15107850]
2. Innan H, Kondrashov F. The evolution of gene duplications: Classifying and distinguishing between models. *Nat Rev Genet*. 2010; 11:97–108. [PubMed: 20051986]
3. Ohno, S. *Evolution by Gene Duplication*. Allen & Unwin; Springer-Verlag; London, New York: 1970.
4. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science*. 2000; 290:1151–1155. [PubMed: 11073452]
5. Lan X, Pritchard JK. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science*. 2016; 352:1009–1013. [PubMed: 27199432]
6. Wapinski I, Pfeffer A, Friedman N, Regev A. Natural history and evolutionary principles of gene duplication in fungi. *Nature*. 2007; 449:54–61. [PubMed: 17805289]



7. Force A, et al. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. 1999; 151:1531–1545. [PubMed: 10101175]
8. Wheeler JC, Shigesada K, Gergen JP, Ito Y. Mechanisms of transcriptional regulation by Runt domain proteins. *Semin Cell Dev Biol*. 2000; 11:369–375. 369–375. [PubMed: 11105901]
9. Levanon D. Spatial and temporal expression pattern of Runx3 (Aml2) and Runx1 (Aml1) indicates non-redundant functions during mouse embryogenesis. *Mech Dev*. 2001; 109:413–417. [PubMed: 11731260]
10. Goyama S, et al. The transcriptionally active form of AML1 is required for hematopoietic rescue of the AML1-deficient embryonic para-aortic splanchnopleural (P-Sp) region. *Blood*. 2004; 104:3558–3564. [PubMed: 15271791]
11. Weirauch MT, et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*. 2014; 158:1431–1443. [PubMed: 25215497]
12. Berg OG, von Hippel PH. Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. *J Mol Biol*. 1987; 193:723–750. [PubMed: 3612791]
13. Hentsch B, Mouzaki A, Pfeuffer I, Rungger D, Serfling E. The weak, fine-tuned binding of ubiquitous transcription factors to the Il-2 enhancer contributes to its T cell-restricted activity. *Nucleic Acids Res*. 1992; 20:2657–2665. [PubMed: 1614851]
14. Jiang J, Levine M. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell*. 1993; 72:741–752. [PubMed: 8453668]
15. Gaudet J, Mango SE. Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. *Science*. 2002; 295:821–825. [PubMed: 11823633]
16. Scardigli R, Baumer N, Gruss P, Guillemot F, Le Roux I. Direct and concentration-dependent regulation of the proneural gene Neurogenin2 by Pax6. *Development*. 2003; 130:3269–3281. [PubMed: 12783797]
17. Tanay A. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res*. 2006; 16:962–972. [PubMed: 16809671]
18. Rowan S, et al. Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. *Genes Dev*. 2010; 24:980–985. [PubMed: 20413611]
19. Crocker J, et al. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell*. 2015; 160:191–203. [PubMed: 25557079]
20. Farley EK, et al. Suboptimization of developmental enhancers. *Science*. 2015; 350:325–328. [PubMed: 26472909]
21. Cannavò E. Genetic variants regulating expression levels and isoform diversity during embryogenesis. *Nature*. 2017; 541:402–406. [PubMed: 28024300]
22. Crocker J, Noon EP, Stern DL. The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution. *Curr Top Dev Biol*. 2016; 117:455–469. [PubMed: 26969995]
23. Nuzhdin SV, Rychkova A, Hahn MW. The strength of transcription-factor binding modulates co-variation in transcriptional networks. *Trends Genet*. 2010; 26:51–53. [PubMed: 20080313]
24. Bravo J, Li Z, Speck NA, Warren AJ. The leukemia-associated AML1 (Runx1)-CBF beta complex functions as a DNA-induced molecular clamp. *Nat Struct Biol*. 2001; 8:371–378. [PubMed: 11276260]
25. Tahirov TH, et al. Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBFbeta. *Cell*. 2001; 104:755–767. [PubMed: 11257229]
26. Komori T, et al. Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell*. 1997; 89:755–764. [PubMed: 9182763]
27. Otto F, et al. Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell*. 1997; 89:765–771. [PubMed: 9182764]
28. Levanon D, et al. The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons. *EMBO J*. 2002; 21:3454–3463. [PubMed: 12093746]

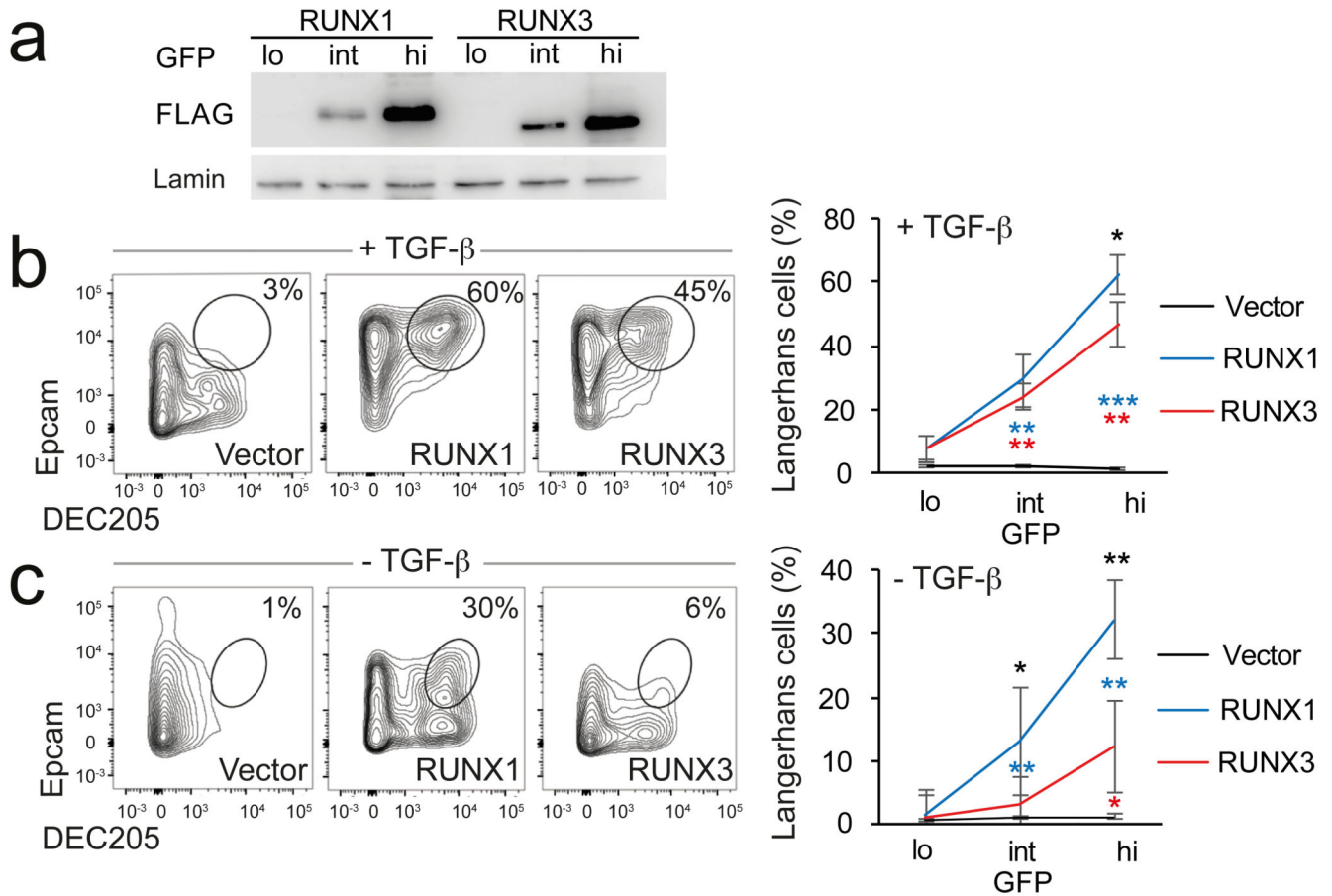
29. Okuda T, van Deursen J, Hiebert SW, Grosveld G, Downing JR. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*. 1996; 84:321–330. [PubMed: 8565077]
30. Wang Q, et al. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc Natl Acad Sci USA*. 1996; 93:3444–3449. [PubMed: 8622955]
31. Collins A, Littman DR, Taniuchi I. RUNX proteins in transcription factor networks that regulate T-cell lineage choice. *Nat Rev Immunol*. 2009; 9:106–115. [PubMed: 19165227]
32. Voon D, Hor YT, Ito Y. The RUNX complex: reaching beyond haematopoiesis into immunity. *Front Immunology*. 2015; 146:523–36.
33. Bae SC, Ito Y. Regulation mechanisms for the heterodimeric transcription factor, PEBP2/CBF. *Histol Histopathol*. 1999; 14:1213–1221. [PubMed: 10506937]
34. Fainaru O, et al. Runx3 regulates mouse TGF-beta-mediated dendritic cell function and its absence results in airway inflammation. *EMBO J*. 2004; 23:969–979. [PubMed: 14765120]
35. Chopin M, et al. Langerhans cells are generated by two distinct PU.1-dependent transcriptional networks. *J Exp Med*. 2013; 210:2967–80. [PubMed: 24249112]
36. Borkowski TA, Letterio JJ, Farr AG, Udey MC. A role for endogenous transforming growth factor beta 1 in Langerhans cell biology: the skin of transforming growth factor beta 1 null mice is devoid of epidermal Langerhans cells. *J Exp Med*. 1996; 184:2417–2422. [PubMed: 8976197]
37. Bruno L, et al. Runx proteins regulate Foxp3 expression. *J Exp Med*. 2009; 206:2329–2337. [PubMed: 19841090]
38. Kitho A, et al. Indispensable role of the Runx1-Cbfbeta transcription complex for in vivo-suppressive function of FoxP3+ regulatory T cells. *Immunity*. 2009; 31:609–620. [PubMed: 19800266]
39. Rudra D, et al. Runx-CBFBeta complexes control expression of the transcription factor Foxp3 in regulatory T cells. *Nat Immunol*. 2009; 10:1170–1177. [PubMed: 19767756]
40. Josefowicz SZ, Lu LF, Rudensky AY. Regulatory T cells: mechanisms of differentiation and function. *Annu Rev Immunol*. 2012; 30:531–564. [PubMed: 22224781]
41. Taniuchi I, et al. Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. *Cell*. 2002; 111:621–633. [PubMed: 12464175]
42. Telfer JC, Hedblom EE, Anderson MK, Laurent MN, Rothenberg EV. Localization of the domains in Runx transcription factors required for the repression of CD4 in thymocytes. *J Immunol*. 2004; 172:4359–4370. [PubMed: 15034051]
43. Berger MF, et al. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotech*. 2006; 24:1429–1435.
44. Boucheron N, et al. CD4(+) T cell lineage integrity is controlled by the histone deacetylases HDAC1 and HDAC2. *Nat Immunol*. 2014; 15:439–448. [PubMed: 24681565]
45. Aken BL, et al. Ensembl 2017. *Nucleic Acids Res*. 2017; 45:635–642.
46. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007; 24:1586–1591. [PubMed: 17483113]
47. Nah GS, Lim ZW, Tay BH, Osato M, Venkatesh B. Runx family genes in a cartilaginous fish, the elephant shark (*Callorhynchus milii*). *PLoS One*. 2014; 9:e93816. [PubMed: 24699678]
48. Takahashi S, et al. GATA factor transgenes under GATA-1 locus control rescue germline GATA-1 mutant deficiencies. *Blood*. 2000; 96:910–916. [PubMed: 10910904]
49. Cruz-Guilloty F, et al. Runx3 and T-box proteins cooperate to establish the transcriptional program of effector CTLs. *J Exp Med*. 2009; 206:51–59. [PubMed: 19139168]
50. Yan J, Liu Y, Lukasik SM, Speck NA, Bushweller JH. CBFbeta allosterically regulates the Runx1 Runt domain via a dynamic conformational equilibrium. *Nat Struct Mol Biol*. 2004; 11:901–906. [PubMed: 15322525]



**Fig. 1. RUNX paralog evolution and RUNT domain conservation**

a) Origin of RUNX paralogs (see methods).

b) RUNX sequence alignment. Sequences are coloured in blue gradient, according to their BLOSUM conservation score relative to the average residue. The bottom track highlights the global conservation.

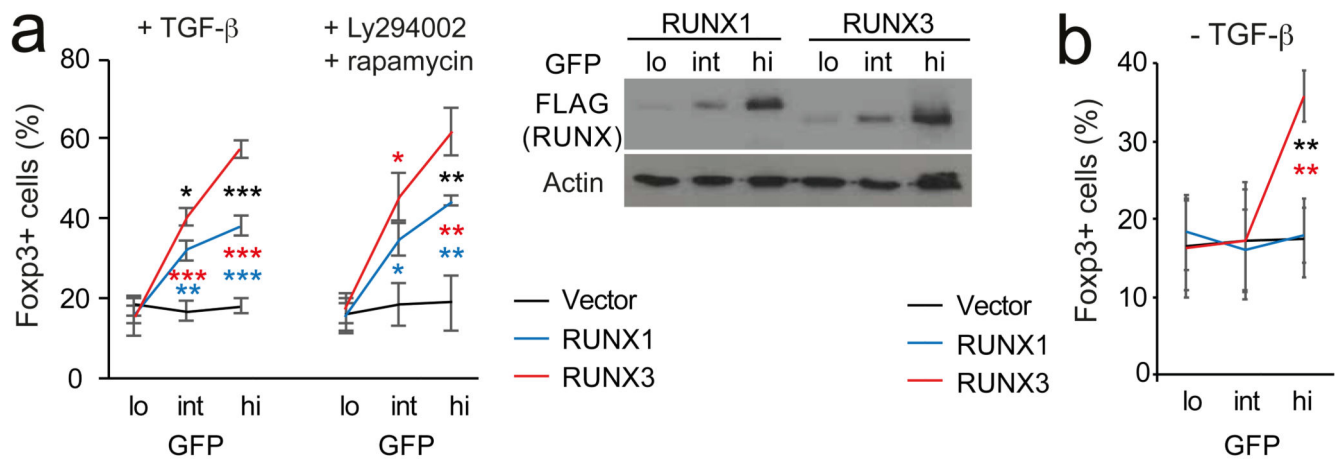


**Fig. 2. Selective deployment of RUNX paralogs enables signal-responsive Langerhans cell differentiation.**

a) Expression of FLAG-RUNX by immature DC transduced with FLAG-RUNX1- or -3-IRES-GFP and sorted into GFP-low, intermediate, and high (lo/int/hi). Levels of retrovirally encoded RUNX3 protein were comparable to endogenous RUNX3 protein in wild-type immature DC (Supplementary Fig. 1a, right). Lamin is the loading control. Representative of 3 independent experiments.

b) *Spil*-deficient BM cells were cultured with GM-CSF and TGF- $\beta$ , and transduced with RUNX-IRES-GFP or IRES-GFP control vector after 24h. Numbers indicate percentages of CD11c<sup>+</sup> MHC class II<sup>+</sup> DEC205<sup>+</sup> Epcam<sup>+</sup> LH cells after 72h. Right: mean  $\pm$  SD of 3 independent experiments. See Supplementary Fig. 1b for numbers of cells recovered. \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$  by two-tailed T test between Runx1 and -3 (black), Runx1 and control vector (blue), Runx3 and control vector (red).

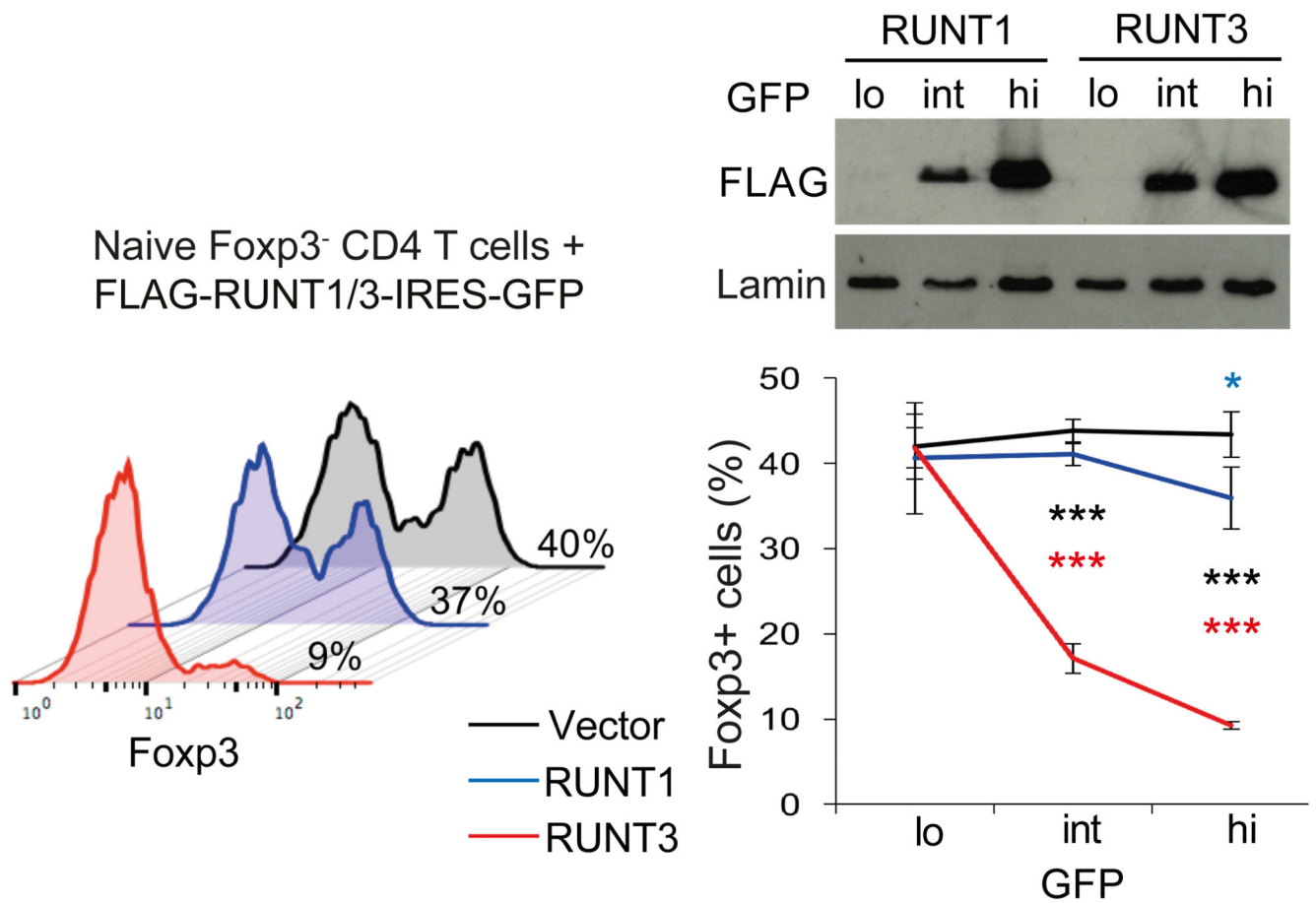
c) *Spil*-deficient immature DC were cultured without TGF $\beta$  and analysed as in b. Right: mean  $\pm$  SD of 3 independent experiments. Statistics as in b.



**Fig. 3. Selective deployment of RUNX paralogs enables signal-responsive induction of the Treg cell signature transcription factor Foxp3.**

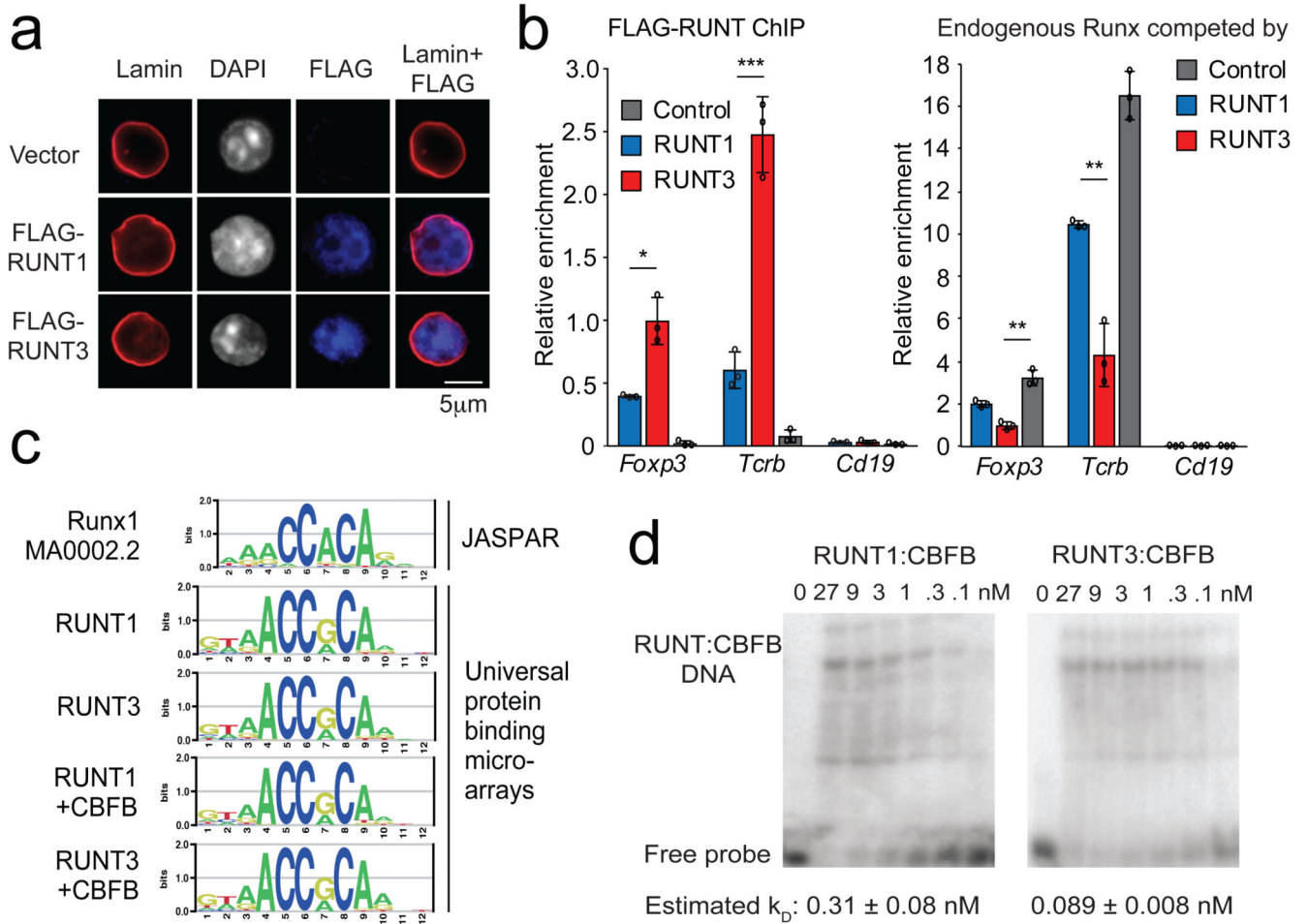
a) Intracellular staining (left) shows rescue of Foxp3 induction in *Runx1*-deficient naive CD4 T cells at equivalent expression of RUNX1 and -3 (right). Mean  $\pm$  SD of 3 independent experiments. Levels of retrovirally encoded RUNX1 protein were comparable to endogenous RUNX1 in wild-type CD4 T cells (Supplementary Fig. 2a). See Supplementary Fig. 2b for numbers of cells recovered from this and subsequent T cell experiments. \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$  by two-tailed T test between Runx1 and -3 (black), Runx1 and control vector (blue), Runx3 and control vector (red).

b) *Runx1*-wild type naive CD4 T cells were activated for 18h, transduced with RUNX1- or -3-IRES-GFP or control vector and cultured without TCR stimulation, without TGF- $\beta$  or PI3K/mTOR inhibitors. Mean  $\pm$  SD of 3 independent experiments. Statistics as in a.



**Fig. 4. Functional differences in *Foxp3* regulation encoded by the RUNT domain**

Expression of FLAG-RUNT in GFP<sup>lo/int/hi</sup> CD4 T cells transduced with RUNT1 or -3 (top). Naive CD4 T cells were activated, transduced with RUNT1/3 or control vector, and cultured with TGF- $\beta$ . Percentages of Foxp3<sup>+</sup> cells (left). Mean  $\pm$  SD of 3 independent experiments (right). \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$  by two-tailed T test between RUNT1 and -3 (black), RUNT1 and control vector (blue), RUNT3 and control vector (red).



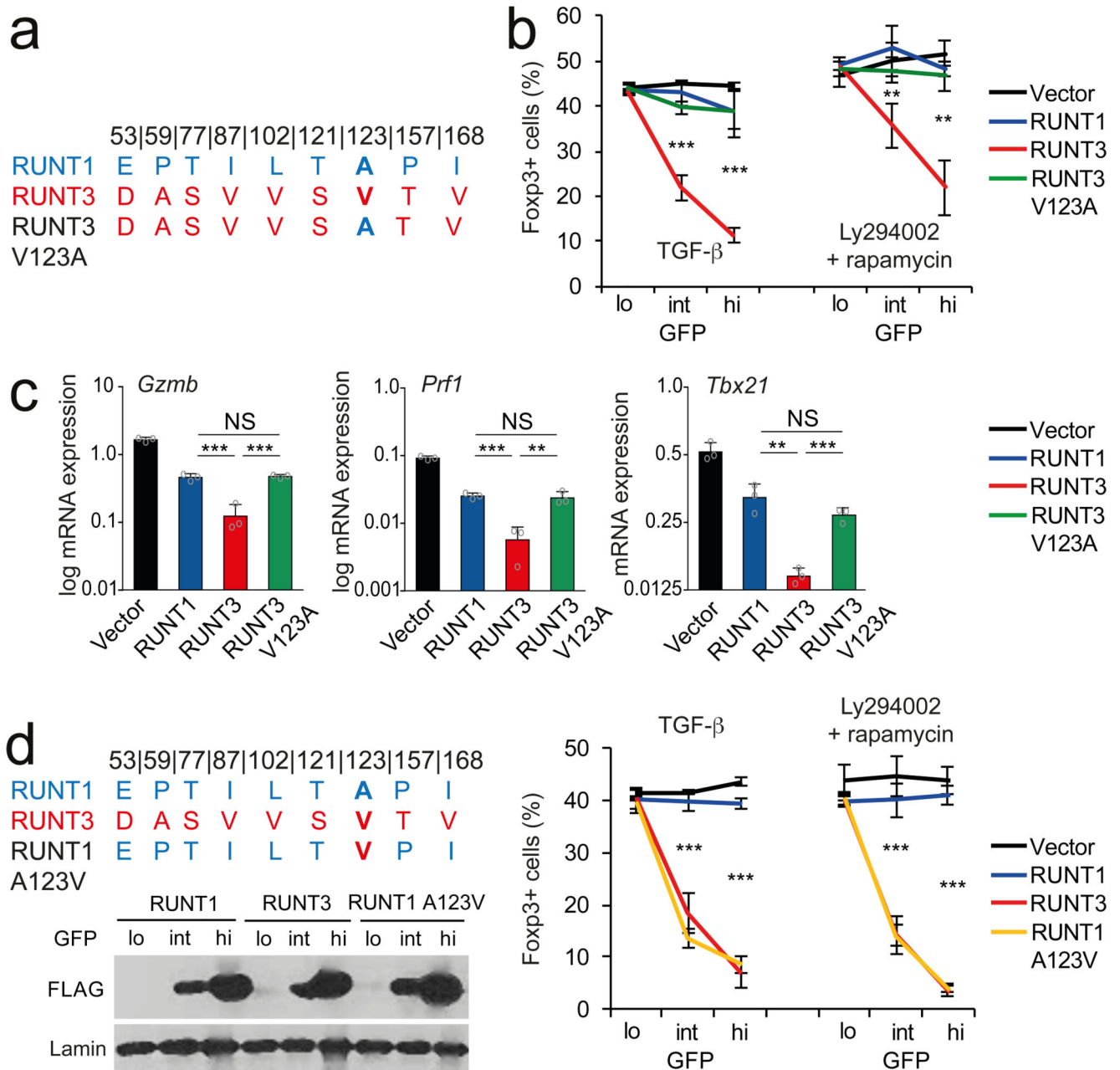
**Fig. 5. The Runx1 and Runx3 RUNT domains have different DNA binding affinities**

a) Nuclear localization of FLAG-tagged RUNT domains did not require the putative nuclear localisation sequence immediately downstream of the RUNT domain<sup>42</sup> or addition of exogenous nuclear localisation sequences. See Supplementary Fig. 3a for RUNT domain constructs used. Representative of 2 independent experiments.

b) Left: FLAG ChIP-PCR detects chromatin binding of FLAG-tagged RUNT domains to canonical RUNX binding sites in the *Tcrb* enhancer and the *Foxp3* promoter. *Cd19* is a negative control. Three biological replicates. \*  $P < 0.05$ . Right: ChIP of endogenous RUNX proteins using an antibody (Abcam ab 92336) against an epitope outside the RUNT domain in the C-terminal domain of Runx1, 2 and 3. *Foxp3*: *Foxp3* CNS2, *Tcrb*: *Tcrb* enhancer, *Cd19*: negative control. Mean  $\pm$  SD of 3 independent experiments. \*  $P < 0.01$ , \*\*  $P < 0.005$ , \*\*\*  $P < 0.001$  by two-tailed Student's T test.

c) Canonical RUNX motif (top) and sequence motifs derived by universal protein binding microarrays for RUNX and RUNX:CBF1 complexes (bottom).

d) Titration EMSAs of RUNT1, RUNT3 complexed with CBF1 binding to a canonical RUNX binding site in the *Foxp3* promoter. CBF1 protein was present in excess (270 nM). See Supplementary Fig. 4 for probe sequence, replicates, and estimates of dissociation constants. Representative of 3 independent experiments.



**Fig. 6. Identification of residues that functionally distinguish paralogous RUNT domains**

a) RUNT1 and -3-specific amino acids, numbers refer to position in mouse RUNX1. With the exception of I168, none of these contact DNA.

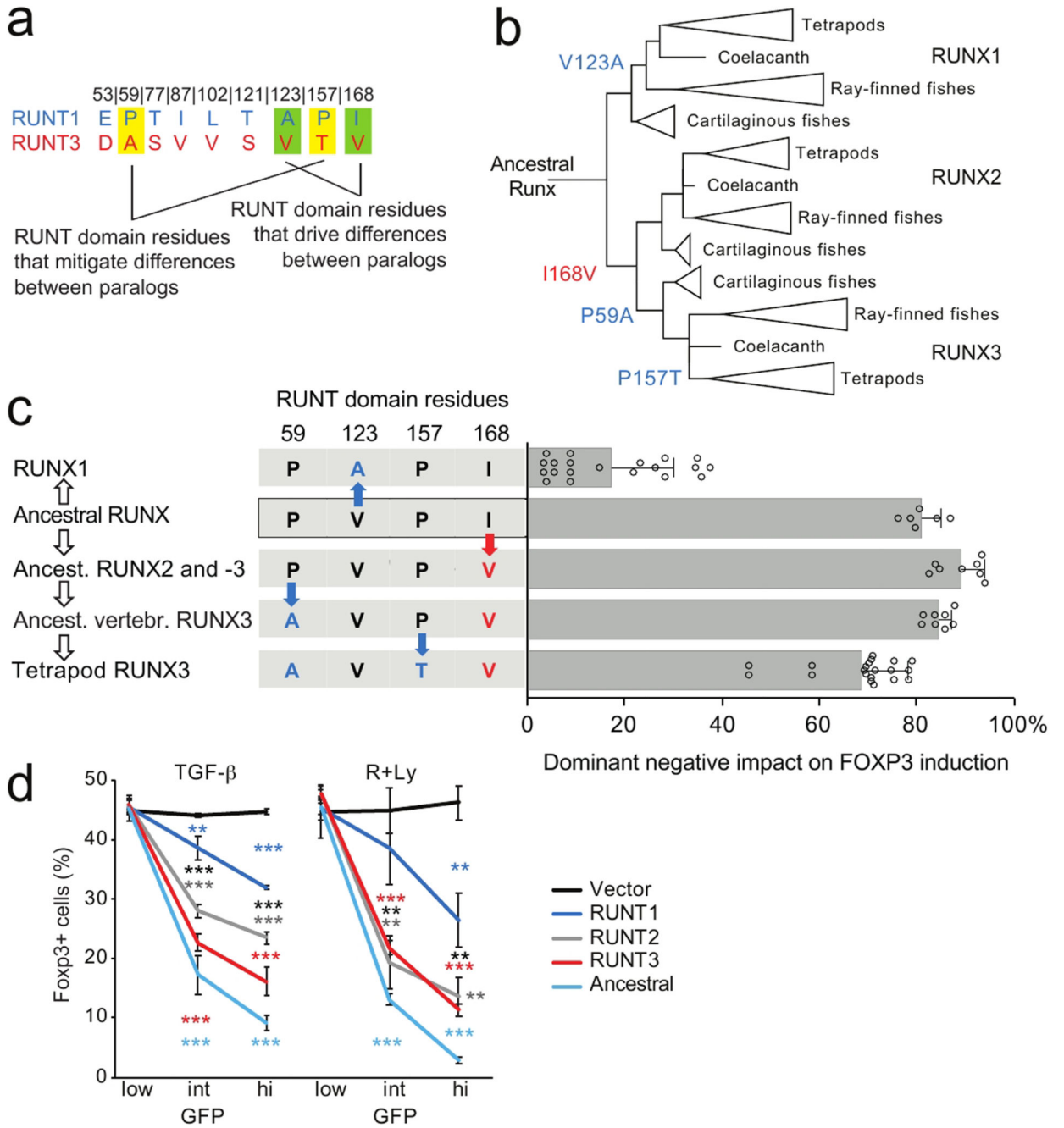
b) Replacement of RUNT3 V123 by the RUNT1 A123 (V123A) weakens the dominant negative activity of RUNT3. Mean  $\pm$  SD of 3 independent experiments. \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$  by two-tailed T test between RUNT3 and RUNT3 V123A.

c) V123 affects the regulation of the RUNX target genes *Gzmb*, *Prf1*, and *Tbx21*. Mean  $\pm$  SD of 3 independent experiments. Results for all RUNT domain constructs were



significantly different from control vector ( $P < 0.001$  by two-tailed Student's T test). \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$  by two-tailed 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). Mean  $\pm$  SD of 3 independent experiments. \*\*\*  $P < 0.001$  by two-tailed T test between RUNT1 and RUNT1 A123V.



**Fig. 7. Runx domain evolution towards sub-maximal strength.**

- a) Amino acid residues that specify functional differences between RUNX1 and -3.
- b) Evolution of RUNT domains. Residues 59, 123, 157 and 168 are highlighted. See text for details. The reconstructed ancestral RUNT domain features residues P59, V123, P157 and I168.
- c) Functional evolution of RUNT domain residues. Red and blue letters and arrows indicate amino acid substitutions that increase or reduce RUNT domain activity (n = 6 to 20, mean  $\pm$  SD of 3 to 10 independent experiments). The X axis shows the dominant negative impact of

each construct on Foxp3 induction: at 0% dominant negative activity, the number of Foxp3-expressing cells is the same as for control vector, at 100% dominant negative activity there are no Foxp3-expressing cells. Constructs to test the activity of the ancestor of Runx2 and -3 and early vertebrate Runx3 contained P59, L102 and T121, which in the context of the ancestral RUNT domain were functionally equivalent to E53, L102, and T121 (Supplementary Fig. 8a). See Supplementary Fig. 3a for sequences. Two-tailed Student's T test RUNT3 versus early vertebrate Runx3:  $P=9.52 \times 10^{-5}$ , RUNT3 versus ancestor of Runx2 and -3  $P=4.84 \times 10^{-6}$ , RUNT3 versus RUNT domain of putative ancestral Runx:  $P=0.0055$ , RUNT3 versus RUNT1  $P=4.89 \times 10^{-17}$ .

d) Functional comparison of the putative ancestral RUNT domain with the RUNT domains of present-day RUNX1, -2 and -3. Mean  $\pm$  SD of 3 independent experiments \*  $P<0.05$ , \*\*  $P<0.01$ , \*\*\*  $P<0.001$  by two-tailed T test between RUNT1 and -3 (black), RUNT1 and control vector (blue), RUNT3 and control vector (red).