Rora regulates activated T helper cells during inflammation

Authors: Liora Haim-Vilmovsky[1,2,♣], Jennifer A Walker[3,♣], Johan Henriksson[4,14,♣], Zhichao Miao[1,2], Eviatar Natan[5], Gozde Kar[1,2], Simon Clare[2], Jillian L Barlow[3], Evelina Charidemou[6], Lira Mamanova[2], Xi Chen[2,10], Valentina Proserpio[7], Jhuma Pramanik[2,8], Steven Woodhouse[9], Anna V Protasio[15], Mirjana Efremova[2], Matt Berriman[2], Gordon Dougan[2], Jasmin Fisher[11], John C Marioni[1,2,12], Andrew NJ McKenzie[3], Sarah A Teichmann[1,2,13*]

Affiliations:

[1]EMBL-European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK
[2]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom
[3]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, UK
[4]Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden
[5]The Aleph Lab, Oxford, UK
[6]Department of Biochemistry, University of Cambridge, UK
[7]Department of Life Sciences and System Biology, University of Turin and Italian Institute for Genomic Medicine-IIGM, Turin
[8]Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK
[9]Department of Genetics, University of Pennsylvania, US
[10]Department of Biology, South University of Science and Technology of China, Shenzhen, China
[11]UCL Cancer Institute, University College London, London, UK
[12]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK
[13]Theory of Condensed Matter, Cavendish Laboratory, 19 JJ Thomson Ave, Cambridge CB3 0HE, UK
[14]Department of Molecular Biology, Umeå University, Sweden
[15]Cambridge Institute for Medical Research, University of Cambridge, CB2 0XL, UK

♣ Equal contribution
* To whom correspondence should be addressed: st9@sanger.ac.uk

keywords: CD4 T helper cell, mouse, Rora, RNA-seq, type 2 immunity, cytokine, screen, FACS, mass spectroscopy, KO, overexpression, lung, spleen, small intestine lamina propria, lymph node, screen, cytokine, in vitro, in vivo

Anna Protasio: avp25@cam.ac.uk	Mirjana Efremova: me5@sanger.ac.uk
Liora Haim-Vilmovsky: liora.haim@gmail.com	Simon Clare: sc7@sanger.ac.uk
Jennifer A Walker: jwalker@mrc-lmb.cam.ac.uk	Lira Mamanova: lm4@sanger.ac.uk
Zhichao Miao: zmiao@ebi.ac.uk	Xi Chen: chenx9@sustech.edu.cn
Evelina Charidemou: ec585@cam.ac.uk	Matt Berriman: mb4@sanger.ac.uk
Johan Henriksson: mahogny@areta.org	Gordon Dougan: gd1@sanger.ac.uk
Eviatar Natan: eviatarhj@gmail.com	Jasmin Fisher: jasmin.fisher@ucl.ac.uk
Gozde Kar: gk12@sanger.ac.uk	John Marioni: marioni@ebi.ac.uk
Steven Woodhouse: Steven.Woodhouse@pennmedicine.upenn.edu	Andrew NJ McKenzie: anm@mrc-lmb.cam.ac.uk
Jillian L Barlow: jbarlow@mrc-lmb.cam.ac.uk	Valentina Proserpio: valentina.proserpio@unito.it
Jhuma Pramanik: jp19@sanger.ac.uk	Sarah A Teichmann: st9@sanger.ac.uk

Abstract (150w)

The transcription factor *Rora* has been shown to be important for the development of ILC2 and the regulation of ILC3, macrophages and Treg cells. Here we investigate the role of *Rora* across CD4+ T cells, both *in vitro* as well as in the context of several *in vivo* type 2 infection models. We dissect the function of *Rora* using overexpression and a CD4-conditional *Rora*-knockout mouse, as well as a RORA-reporter mouse. We establish the importance of *Rora* in CD4+ T cells for controlling lung inflammation induced by *Nippostrongylus brasiliensis* infection, and have measured the effect on downstream genes using RNA-seq. Using a systematic stimulation screen of CD4+ T cells, coupled with RNA-seq, we identify upstream regulators of *Rora*, most importantly IL-33 and CCL7. Our data suggest that *Rora* is a negative regulator of the immune system, involved in a range of pathways, under control of the local microenvironment.

Introduction (500w)

A number of transcription factors are thought to control the overall function of T helper cells, such as *Tbx21* for T helper type 1 (Th1) and *Gata3* for Th2. In this study we focus on the gene Retinoic acid receptor-related orphan receptor alpha (*Rora*), a nuclear receptor with similarities to the Th17 regulator *Rorc*[1]. To date, *Rora* is known to be important for the development of type 2 innate immune cell (ILC2)[2,3] and for cytokine production by ILC3[4]. In liver, it controls inflammation by promoting the macrophage M2 polarization[5]. Recently it was shown to have an important function for skin Tregs[6], where it is required for proper regulation of the immune response in atopic dermatitis. Given this, we hypothesised that *Rora* might be broadly involved in regulating an inflammatory response as part of a common core immunoregulatory program. To investigate this we assessed the function of *Rora* in 4 different infection models involving, but not limited to, type 2 immunity (Fig. 1).

By employing bulk RNA sequencing (RNA-seq), single cell RNA-seq (scRNA-seq) and mass spectrometry (MS) analysis of T helper cells, we have investigated the role played by *Rora* in the context of *in vitro* culture and infection models. Our ur data suggest that Rora is expressed by activated $CD4^+$ T cells and its expression correlates with effector function, such as the expression of Th lineage-defining transcription factors and the production of cytokines.

Results

Rora is broadly expressed in activated T helper cells *in vivo*

To get an overview of *Rora* expression in T helper cells during worm infection, mice were infected with *Nippostrongylus brasiliensis* (*N. brasiliensis*), a tissue migrating parasitic roundworm of rodents. *N. brasiliensis* initiates infection at the skin, the site of invasion, and then migrates via the circulatory system

to the lungs. Through exploitation of host lung clearance, worms are transported to the intestine, from where they are expelled. In total, this infection is resolved within 7-10 days (Fig. 1).

To get an overview of the infection at the molecular level, we performed single cell quantitative reverse transcription PCR (scRT-qPCR, Biomark). We prepared cells from the lungs, small intestine lamina propria (siLP), mediastinal lymph nodes (medLN) and mesenteric lymph nodes (mesLN) of mice, 3, 5, and 7 days after *N. brasiliensis* infection (Fig. 2a). Tissues were also collected from uninfected mice as controls. Both CD3+CD4+SELL+ (naive) and CD3+CD4+SELL- (activated) cells were sorted. We examined 440 individual CD3+CD4+ T cells for the expression of 96 genes including housekeeping genes, different immune cell markers, T helper markers, cytokines, cytokine receptors, chemokine receptors, known and candidate regulators for Th cell differentiation (Fig. 2b; Supplementary S2). Consistent with a predominantly type-2 immune response, 70% of the CD3+CD4+SELL- cells expressed *Gata3*, whilst 26% expressed *Foxp3* and only 2% contained *Tbx21* or *Rorc*. Interestingly, in this population, we noted a significant overlap of *Gata3* and *Rora* ($p=4*10^{-4}$), and also of *Rora* and *Foxp3* ($p=2*10^{-4}$) (Fig. 2b).

To investigate *Rora* activity in activated Th cells, we examined *Rora* expression using a $Rora^{+/teal}$ reporter mouse, in which the gene encoding Teal fluorescent protein is targeted to the 3' end of the *Rora* coding sequence and expressed from a self-cleaving T2A site (submitted). Flow cytometric analysis of reporter gene expression indicated that *Rora* was expressed in a proportion of $CD4^+$ T cells isolated from MedLN, spleen, lung and siLP (Fig. 2c). We noted that all *Rora* expressing cells possessed an activated phenotype, as evidenced by their expression of CD44, ICOS and CD38 and down-regulation of SELL (Fig. 2d). Accordingly, $Rora^+$ T cells were therefore significantly enriched within the activated T cell subset, comprising 40-60% of this population, but absent from the pool of naive cells (Fig. 2e). This correlation of *Rora* expression with T cell activation is in contrast with RNAseq analysis made during the first 72h of *in vitro* culture of Th0 and Th2 cells, as previously reported (Fig. 2f)[7]. There, in both mouse and human, *Rora* is expressed in naive cells, with a sudden increase around 2-4 hours of activation, before a gradual decrease in expression (overall slightly higher expression in Th2 over Th0). This agrees with another *in vitro* activation dataset where the *Rora* expression level is unchanged and low, in both Th17 and Th1 cells, following activation[8]. To compare with the *in vivo* situation, we generated RNA-seq data from CD4+ T cells, isolated from the spleens of *N. brasiliensis* infected wild-type mice 1, 3 and 10 days after infection, and overlapped with the *in vitro* data (Fig. 2f). Here *Rora* is seen to increase, reaching a high and fairly sustained level around day 3. This is in agreement with a previous single-cell analysis of Th1/Tfh cells during malaria infection where *Rora* expression is seen to increase over time[9] (Sup. Fig. 7), as well as a study on atopic dermatitis[10]. Taken together, as *Rora* is consistently present *in vivo* but not *in vitro*, these 6 datasets suggest that *Rora* is induced not primarily by activation, but by an external cue from the micro-environment.

Building upon this observation, we examined our scRT-qPCR data across different organs (Fig. 2b). Overall, the expression of *Rora* was relatively constant in CD3+CD4+ cells 3-7 days after infection (Fig. 2f). However, in terms of tissue, it was expressed more frequently in the siLP and lungs, where 37% of cells expressed *Rora*, compared to 20% of the cells in lymph nodes. This pattern was confirmed using flow cytometry, where 10.9±1.9% of CD4+ cells in the lung were RORA+, 10.1±1.1% from the spleen,

27.9±5.8% from the small intestine lamina propria and only 4.0±0.6% from medLN (Fig. 2d). This tissue distribution likely reflects the relative frequency of activated Th cells in each location.

Rora has so far been seen in activated cells, partly overlapping with primary markers of Th2 and Treg (*Gata3* and *Foxp3*). To fully confirm the presence of *Rora* among different T helper cell types, as defined by the full transcriptional program, we used scRNA-seq on lymphoid and non-lymphoid tissue, isolating CD3+CD4+SELL- cells from the lungs 30 days after infection (Fig. 2h). As reference points we included CD3+CD4+Rora-TEAL+ and CD3+CD4+Rora-TEAL- cells from *N. brasiliensis* infected $Rora^{+/teal}$ reporter mice. Lungs, spleens, medLN and mesLN were also taken from infected and uninfected $Rora^{+/teal}$ mice, 7 days after infection. The cells were clustered and each cluster annotated using common markers (*Sell, Ifng, Gata3, Foxp3*) as Naive, Th1, Th2 and Treg cells (Fig. 2i, all cells are shown in Supplementary Fig. 1). The distribution of *Rora* expression level across the different CD4+ T cell types was investigated (Fig. 2j). Again, we note that *Rora* is expressed broadly among all activated T cells (including Th1).

To confirm that this distribution of Rora expression is not specific to the response to *N. brasiliensis* infection we compared it to another worm model, *Schistosoma mansoni* (*S. mansoni*). CD3+CD4+SELL- cells were collected 6 weeks after infection from spleens and mesLN (Sup. Fig. 8). Single cell RNA-seq was performed and analysis shows that *Rora* is expressed in 29% of the activated cells. Here again, the approximately 30% cells that expressed *Rora* also expressed other key regulators (*Gata3*, *Foxp3*, *Tbx21*).

To conclude, we have performed FACS, scRT-qPCR and scRNA-seq, and confirmed that *Rora* is generally expressed in a proportion of activated T cells during *N. brasiliensis* infection. *Rora* induction is likely to depend on a local environmental cue not present in the usual CD3/CD28 *in vitro* activation system. Thus an *in vivo* system is required to study the physiological role of *Rora.*

Rora is expressed in activated T helper cells during immune responses

To further corroborate these findings, we also investigated a non-worm immunity model, Ragweed pollen (RWP), a common allergen which causes lung inflammation[11]. The $Rora^{+/teal}$ reporter mice were exposed to intranasal RWP or PBS (Phosphate buffered saline, a negative control administration), administered every 2-3 days for 2 or 6 weeks (Fig. 3a). As anticipated, RWP exposure resulted in increased CD4+ T cell counts in the lung and mediastinal lymph node (medLN) after 2 weeks as compared to PBS controls, and this was maintained after 6 weeks of continuous RWP administration (Fig. 3a). In lung, but not medLN, this increase in CD4+ T cell frequency was accompanied by an elevated proportion of T cells expressing *Rora*, increasing from 30% to 40% between 2 and 6 weeks of RWP administration (Fig. 3b and Supplementary figure 2). Consistent with our findings in untreated mice, RORA+ T cells elicited by RWP stimulation had an activated phenotype, characterised by a CD44+SELL- surface profile. (Supplementary Fig. 2). Notably, the RORA+ T cell subset was enriched for cells that had acquired effector function, identified by their expression of the cytokines IFNg, IL-13 and IL-17A, or the Treg-associated transcription factor, FoxP3 (Fig. 3c). Here again we note that *Rora* expressing cells include diverse activated cell phenotypes, that is, Th1, Th2 and Th17.

To understand the involvement of *Rora* in effector cells, we examined its expression in cytokine-expressing cells. From the *N. brasiliensis* model single-cell RT-qPCR, 83% of the cells that expressed either IL-4, IL-13 or IL-10, also expressed *Rora* (Fig. 2b). The corresponding number in the *S. mansoni*-infected mice is 71% (Sup Fig. 8, scRNA-seq)

To investigate whether *Rora* was expressed by Th cells responding to secondary immune challenge, we challenged $Rora^{+/teal}$ and control mice with 2W1S peptide intranasally, in conjunction with the protease allergen papain, which promotes a type-2 immune response (Fig. 3d). After secondary immunisation with 2W1S/papain 3 weeks after the initial stimulation, we were able to identify 2W1S- responsive CD4+ T cells by tetramer staining[12] and FACS analysis. Tetramer+ CD4+ T cells in both the lung and medLN were uniformly CD44+ and Rora-Teal+ (Fig. 3d), indicating that recently activated Th cells express *Rora*. As with the *N. brasiliensis* infection, in the *S. mansoni* model we found significant overlap of *Foxp3* and *Rora* (Sup.Fig 8, $p=2*10^{-4}$). A higher fraction of FOXP3+ cells was also found in sorted RORA+ relative to RORA- cells, in the RWP model (Fig. 3f). The gap is increased when the exposure to the pollen was longer. The same could also be seen after mice were exposed to 2W1S+papain. The RORA+ population contained 20% of the FOXP3+ cells, while the RORA-, only 4% of the cells.

Overall our data show that *Rora* is expressed in Th cells in worm immune models as well as in allergen type 2 immunity. We find *Rora* transcripts present in activated Th cells, not only in Tregs, but also in Th1 and Th2, as well as activated $CD4^+$ T cells responding to a secondary immune challenge. *Rora* is also more correlated with cytokine- releasing cells, which suggests a function in effector cells.

Rora KO affects inflammation severity

To isolate the *in vivo* physiological function of *Rora* in Th cells, we created a conditional knockout mouse line, where exon 4 of the *Rora* gene was deleted only in CD4+ cells ($CD4^{cre}Rora^{fl/fl}$, verified by RNA-seq, Supplementary Fig. 3), resulting in a dysfunctional RORA protein. We noted no change in Th cell number when comparing $CD4^{cre}Rora^{fl/fl}$ to control (Supplementary Fig. 4). We then infected $CD4^{cre}Rora^{fl/fl}$ mice with *N. brasiliensis* and examined lung inflammation after 30 days. An unbiased image analysis algorithm was developed to detect emphysema and score stained lung sections (Fig 4a). Generally, uninfected samples score the lowest for emphysema, as expected (Fig 4b,c). During infection, single-allele *Rora* KO ($CD4^{cre}Rora^{+/fl}$) has a higher score of emphysema, and double-allele KO ($CD4^{cre}Rora^{fl/fl}$) even higher (p=0.018). To ensure that this is not a strain-specific defect we also compared infected WT mice with no-cre KO mice ($Rora^{fl/fl}$), as well as WT with cre KO mice ($CD4^{cre}$), and see no difference.

We conclude that *Rora* expression in CD4+ T cells modestly promotes lung inflammation associated with *N. brasiliensis* infection. This may reflect a role for Rora in regulating the activity of T helper cells, and may also result from the modulation of Treg activity, as has previously been reported for Rora+ Tregs in skin[6].

Rora overexpression affects several key immune regulatory genes

To probe the role of *Rora* in T cells, we next proceeded to identify the molecular targets of *Rora* in Th2 cells by means of retroviral overexpression. Th2 cells were induced *in vitro*, infected the day after

induction, and on day 5, gene expression was compared by RNA-seq (empty viral vector *vs* overexpression vector, Fig. 4d). The *in vitro* differentially expressed (DE) genes were compared to the genes that are differentially expressed between skin Tregs from wt *vs* $CD4^{cre}Rora^{fl/fl}$ mice (Fig. 4f). Consistent with a previous report[6], we see a strong effect on *tumor necrosis factor receptor superfamily member 25* (*Tnfrsf25)*, which has been shown to be required for Th2 effector function, *e.g.*, allergic lung inflammation[13]. The receptors *C-C chemokine receptor type 2* (*Ccr2*), *Ccr5*, *Interleukin 1 receptor, type 2* (*Il1r2*) and *Tnfrsf23* are in agreement, although with lower fold changes. We also find *Ninjurin 2* (*Ninj2*) which has recently been implicated in endothelial inflammation[14].

We also looked at a previous dataset of *Rora* small interfering RNA (siRNA)-treated human Th17 cells[15] and find that *Tnfrsf25* is DE ($p=2.2*10^{-4}$). In another study of Tregs during steady-state in mouse[10], we find that *Tnfrsf25* is higher in colon Tregs than in lymphoid tissue Tregs ($p=2*10^{-2}$). Thus four datasets, in different T helper cell types, confirm *Tnfrsf25* as a downstream gene.

Overall the two *in vitro* systems strengthen the claim of *Rora* as regulator of *Tnfrsf25*. However the expression of *Tnfrsf25* across Th2, Treg and Naive, but not Th1 (Fig. 2h), show that the function of *Rora* is not limited to Treg cells. Combined with data from other cell types, such as macrophages[5], *Rora* can be said to be widely associated with levels of inflammation.

Rora affects multiple pathways *in vivo*

To see if *Rora* has additional functions *in vivo*, we generated bulk RNA-seq from activated (CD62L-/CD44+) CD4+ T cells from control and $CD4^{cre}Rora^{fl/fl}$ mice from lung, spleen, siLP and colon (30 days after *N. brasiliensis* infection), and compared the DE genes between the different tissues, the overexpression DE genes, and the previous Treg Rora KO data. Some of the possible overlaps are shown in Fig. 4g,h. The full list of fold changes are in Supplemental File S2.

We find the transcription factor *Aryl hydrocarbon receptor nuclear translocator-like* (*Arntl*) to be differentially expressed in activated CD4+ T cells from siLP as well as in the overexpression analysis and in Tregs in skin. *Arntl* is involved in the circadian responsiveness, and is expressed in CD4+ T cells[16]. The *Arntl* promoter contains ROR elements (RORE), and was found to be regulated by *Rora* in mice[17]. RORA promotes *Arntl* expression and thereby maintains the circadian rhythm. We find it to be expressed in all cell types (Fig. 2h). *S100 calcium-binding protein A4* (*S100a4*) is another gene which is differentially expressed in lung, colon, and in Tregs in skin. In the single cell dataset, it is expressed mostly in Tregs but generally in Th1/Th2 (Fig. 2h). C-X-C chemokine receptor type 6 (*Cxcr6*) is differentially expressed in siLP, spleen and in Tregs in skin, and might be involved in Th cell activation[18,19]. *Cxcr6* is expressed in all Th cells but mostly in Th2 cells (Fig. 2h).

The sialyltransferase *St6galnac3* is also upregulated. This enzyme transfers sialic acids to glycolipids and glycoproteins. Along with the previously DE gene *Alox8*, we wondered if *Rora* may have any function on lipid metabolism. RORA has already been found to bind cholesterol[20] and is evolutionarily and mechanistically related to the lipid metabolism gene *Peroxisome proliferator-activated receptor gamma* (*Pparg*). We performed lipid LC-MS on *in vivo* $CD4^{cre}Rora^{fl/fl}$ and WT cells from *N. brasiliensis* infected mice, as well as on *in vitro* WT naive and mature Th0/Th1/Th2 cells for reference. The $CD4^{cre}Rora^{fl/fl}$

presented small differences, and little overlap with the 85 lipids (p=0.05) changed *in vitro* the first 6 days. However, all the top DE lipids are phosphocholines or phospoethanolamines, and some phosphocholines can induce a type 2 immune response[21]. All the differentially expressed lipids are listed in Supplementary File S4.

Downregulated genes include Fas Ligand (*Fasl*) which is differentially expressed in activated CD4 T cells from lung and colon, and in Tregs in skin. *Fasl* is involved in CD4 T cell apoptosis[22], and we find it mostly in Th1 cells (Fig. 2h). *Gzma*, which is required for the suppressive function of Tregs during graft versus host disease *in vivo*[23], is also downregulated in activated cells from colon, skin Tregs and *in vitro* Th2 *Rora* overexpression. *Gzma*, as well, is expressed mainly by Th1 (Fig. 2h).

Overall we see that *Rora* affects several genes *in vivo*, that are differentially expressed between different T helper cell types. It is thus likely that *Rora* has a slightly different role in the different Th subsets. The downstream genes do not appear to share a common pathway. It is possible that *Rora* acts as a helper factor, or tissue residence checkpoint, but is not the primary driver, for multiple other functions.

Rora expression is likely extrinsically regulated

A remaining question is which genes regulate *Rora*, and whether regulation is due to the T cell itself or from an external signal. Insight into when *Rora* is induced would provide understanding of its role. As previously seen, it is generally not expressed during *in vitro* activation, but is upregulated during *in vivo* activation. This discrepancy may be explained by an unknown signal from the environment that is present *in vivo* but not *in vitro*. To find candidate upregulators we screened for *in vitro* upregulated genes during 39 treatments, including cytokines and chemokines, with 15 unstimulated controls, in 3 biological replicates. In short, naive cells were taken from wild type mouse spleens, plated on CD3/CD28 coated plates, with addition of test cytokines/chemokines 24 hours later. 5 days after plating the cells were measured by bulk RNA-seq (Fig. 5a).

The stimulants and their effects are shown in Fig. 5b (DE genes in Supplementary File S2). To validate our perturbation assay we looked at known marker genes. Of the genes having the strongest effect we find the expected genes: IL4 for *Gata3* (Th2), Transforming growth factor beta 1 (TGFb1) for *Foxp3* (Treg), IL6 and partially TGFb1 and IL21 for *Rorc* (Th17[24]). We then looked for the effect on *Rora* expression. The most upregulating stimulants are CCL5, IL15 and IL33. CCL5 is chemotactic for T cells, eosinophils, and basophils. CCL5 signaling through CCR3 has been reported to regulate Th2 (IL4+CD4+ T) cellular responses to promote metastasis of luminal breast cancers[25]. Interestingly also CCR5, another CCL5 receptor, was upregulated during overexpression, implicating a potential autocrine role for this factor. IL15 is produced by non-lymphoid cells and important for the survival of several lymphoid subsets[26]. IL15 KO mice are markedly lymphopenic due to decreased proliferation and decreasing homing to peripheral lymph nodes[27]. Increase in IL15 thus fits our observation of *Rora* expression primarily in activated Th cells outside the lymph nodes. IL6 resulted in upregulation of *Rora* , consistent with previous IL6 *in vitro* culture data[8]. IL6 has several roles, beyond generating Th17 cells *in vitro*, both pro- and anti-inflammatory[28]. IL33 also induces *Rora*, and is mainly associated with Th2 or ILC2 cells[29]. This

agrees with our *in vivo* scRNAseq where *Rora* is slightly elevated in Th2 cells compared to other Th subtypes (Fig. 2h).

The most downregulating stimulants are C-C motif chemokine 22 (CCL22) and stromal cell-derived factor 1 (SDF1a). CCL22 is secreted by dendritic cells and macrophages, and is suggested to induce Treg cell infiltration into the pleural space in patients with malignant effusion[30,31]. In addition, its receptor CCR4 is required for CD4 T cell migration to the skin[32]. In our single cell data set, we find *Ccr4* expression mainly in Th2 and Tregs (Fig. 2h). SDF1a (also known as CXCL12), a ligand for the chemokine receptor CXCR4, promotes bone marrow homing for T cells. It has been shown that Naive T cells downregulate CXCR4 to avoid this[33]. A further look at the Human Protein Atlas show the tissues with expression of SDF1a to be particularly high in spleen, cervix/endometrium, and adipose tissue. In our study we compared with lung and gut, which have less *Cxcl12* expressed. This is consistent with our screening data, suggesting SDF1a may be a negative regulator of *Rora*.

To further validate if a treatment induces *Rora*, we also checked if a treatment induces the same downstream genes as in the *Rora* overexpression experiment. To do this we took genes from the *Rora* overexpression with a Log2 fold change > 2, and correlated the fold change with the corresponding fold change in a treatment (Fig. 5c). Many of the most *Rora* influencing treatments also score similarly by this score. Different fold change cut-offs have been tested for reproducibility (data not shown). In particular, IL33 and CCL7 have a high level of agreement, followed by CCL22 (replicate a). Other treatments appear to have a less direct effect on *Rora*.

Based on our and previous data we conclude that *Rora* expression is mainly associated with an inflammatory state. Several cytokines may influence the expression of *Rora*. The effect on other genes can be studied using a similar screening approach we provide a user-friendly interface to the data at the supplementary website http://data.teichlab.org.

Discussion

We have here investigated the impact of *Rora* on different subsets of CD4+ T cells, in several *in vivo* immune models. We show that *Rora* is generally expressed in activated T helper cells during inflammation, including worm infections and allergy. We expand on previous results and suggest that *Rora* plays a role in the regulation of the immune response, and present in multiple T helper cell types - Th1, Th2 and secondary activated cells as well. Our model for *Rora* is pictured in Fig. 6.

Rora deletion did not alter the overall CD4+ T cell numbers, and therefore does not seem to be involved in T cell proliferation. However, higher inflammation in the lungs of $CD4^{cre}Rora^{fl/fl}$ mice, after *N. brasiliensis* infection, indicates a functional role in CD4+ T cells. A similar effect was shown in skin, when *Rora* was deleted in Tregs only[6]. The effect was demonstrated to be through Treg *Tnfrsf25* signaling. *Tnfrsf25* has been shown to be required for Th2 cell function in lung inflammation[13]. Our analysis of the DE of activated Th cells in the lung and colon of worm infected mice, as well as overexpression, and reanalysis of previously published data, confirms the importance of *Tnfrsf25* in Tregs, but that its origin may include other activated T helper cell subsets.

A higher fraction of activated T helper cells (both Th and Tregs) contained *Rora*, when located in non-lymphoid tissues (lung and siLP) compared to lymph node tissues. The same pattern was shown previously for skin-located versus LN-located Tregs[6]. The enrichment of Rora+ cells in peripheral tissues may reflect its correlation with T cell activation and effector function, or alternatively may relate to a role in influencing T cell migration.. T cells migrate in and out of the lymph node over the day[34], and *Rora* is already known to be involved in circadian rhythm. Our data strengthens this claim as the circadian gene *Arntl* is DE in both Tregs and other activated CD4 cells, in agreement with previous studies[17,35]. A recent study on lymphocyte circadian migration identifies the cytokine SDF1a as essential for this process[34]. Interestingly, SDF1a represented a negative regulator of *Rora* in our screen. Upregulation of *Rora* might thus occur once the Th cell has reached its final destination at the infection site or at least in any peripheral tissue, where the cell also start to express cytokines and act as an effector Th cell. However other cytokines in the local microenvironment also affect *Rora* and a comparison on relative signaling strengths is needed to quantitatively model its impact on Th cells.

The previous *Rora* KO sg mice display a range of phenotypes, *e.g.*, reduced body weight gain and neuronal disorders. Assuming the gene regulatory network is preserved in other cell types, our work provides hypothetical mechanisms to explore. Since *Rora* controls PPARa[36], it may be a possible intermediate, with our data suggesting effects on phosphocholines. Alternatively, knowing that the SDF1a-CXCR4 interaction is also important for neuronal guidance, SDF1a-mediated regulation of Rora may also explain any neuronal impact.

To conclude, *Rora* has been found to be important for several immune cell types, in different contexts. It is involved in multiple pathways and appears to link cell migration with cell cycle, but also affects the outcome of inflammation through, but not limited to, *Tnfrsf25*. In its natural context, *Rora* is mainly associated with non-lymphoid tissue and activated T cells. *In vitro* culture suggests that this is due to environmental stimuli. By generating a large systematic transcriptomic characterization of cytokine effects on T helper cells, we have found several cytokines that may explain be responsible for *Rora* induction in other tissues. With the large number of conditions tested in this study we are one step closer to understanding the multiple functions of this gene. All data is available for further browsing at http://data.teichlab.org.

Acknowledgements

We would like to thank Helen Jolin for help with mice, MRC Ares and Biomed staff, and Bee Ling Ng, Chris Hall, Jennie Graham, Maria Daly, Fan Zhang and Martyn Balmont for help with cell sorting. Ayesha Jinat provided technical assistance with RNA-seq.

Author contributions

L.H.V designed the experiments, performed cell isolations, qPCR experiments and analysis, prepared cells for single cell RNA-seq experiments, bulk RNA-seq experiments, adoptive transfer and FACS experiments, did the C1 scRNA-seq experiments, and co-wrote the manuscript. J.W designed

experiments, performed tissue and cell isolations, and most of the flow cytometry experiments and their analysis. J.H helped design some of the experiments, analyzed the data, performed the bulk RNA-seq experiments, analyzed some of the FACS data, did the overexpression and cytokine screen experiments, microscopy and image analysis, and co-wrote the manuscript. Z.M and G.K helped analyze the scRNA-seq data. E.N helped preparing tissues for qPCR experiments. S.W helped analysing scRT-qPCR data. J.L.B helped with the Ragweed model. M.E helped classify NKT cells. L.M helped with the single cell RNA-seq cell classification. S.C did the schistosoma worm infection. X.C and V.P helped generally with the experiments. A.V.P. harvested tissue from schistosoma-infected mice. G.D provided the *Schistosoma mansoni* infection model and advised on the study. J.F helped with the biomark data analysis. J.M helped with and advised on the analysis. A.C performed the MS and corresponding sample preparation. A.M provided most of the infection models, performed the histology scoring, and advised. S.A.T advised and conceived the project.

Funding

L.H.V is supported by EMBO (award number ALTF 698-2012), Directorate-General for Research and Innovation (FP7-PEOPLE-2010-IEF, ThPLAST 274192) and an EMBL Interdisciplinary Postdoctoral fellowship, supported by H2020 Marie Skłodowska Curie Actions. J.H. is funded by the Swedish Research Council, and S.A.T. by the European Research Council grant ThDEFINE. V.P. is supported by FUV - Fondazione Umberto Veronesi. Wellcome Sanger Institute core facilities are supported by grant WT206194. J.W. and A.N.J.M are supported by Wellcome (100963/Z/13/Z) and UK Medical Research Council (U105178805). Z.M. is supported by a Single Cell Gene Expression Atlas grant from the Wellcome Trust (nr. 108437/Z/15/Z).

Competing Interests

None declared

Data and materials availability

The sequencing data has been deposited at ArrayExpress (E-MTAB-7694, E-MTAB-8000, E-MTAB-8001, E-MTAB-8003), and the image data at Zenodo (https://doi.org/10.5281/zenodo.2554078). The R code used for the analysis is available on Github (https://github.com/mahogny/liorarora).

Supplementary files

S1. Biomark single cell RT-qPCR data, raw data and list of TaqMan probes

S2. RNA-seq count data and condition matrices

S3. List of antibodies

S4. LC-MS lipid data

Figures and legends

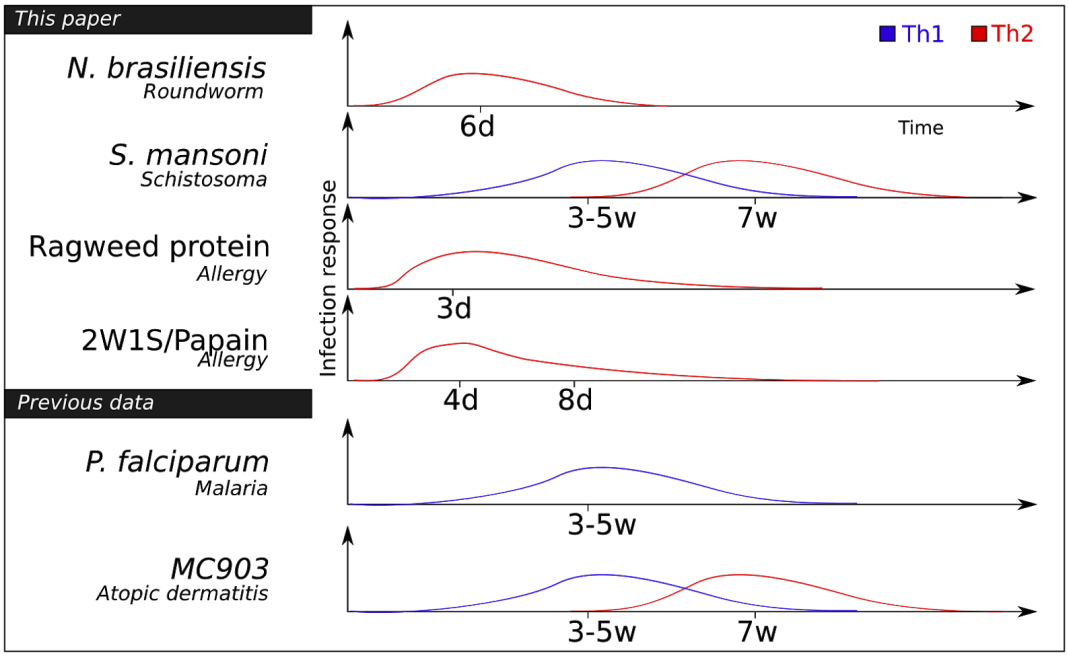

Fig. 1: Overview of the experiments The immune response to the infection models considered in this paper, over time and which T helper cells are involved according to literature.

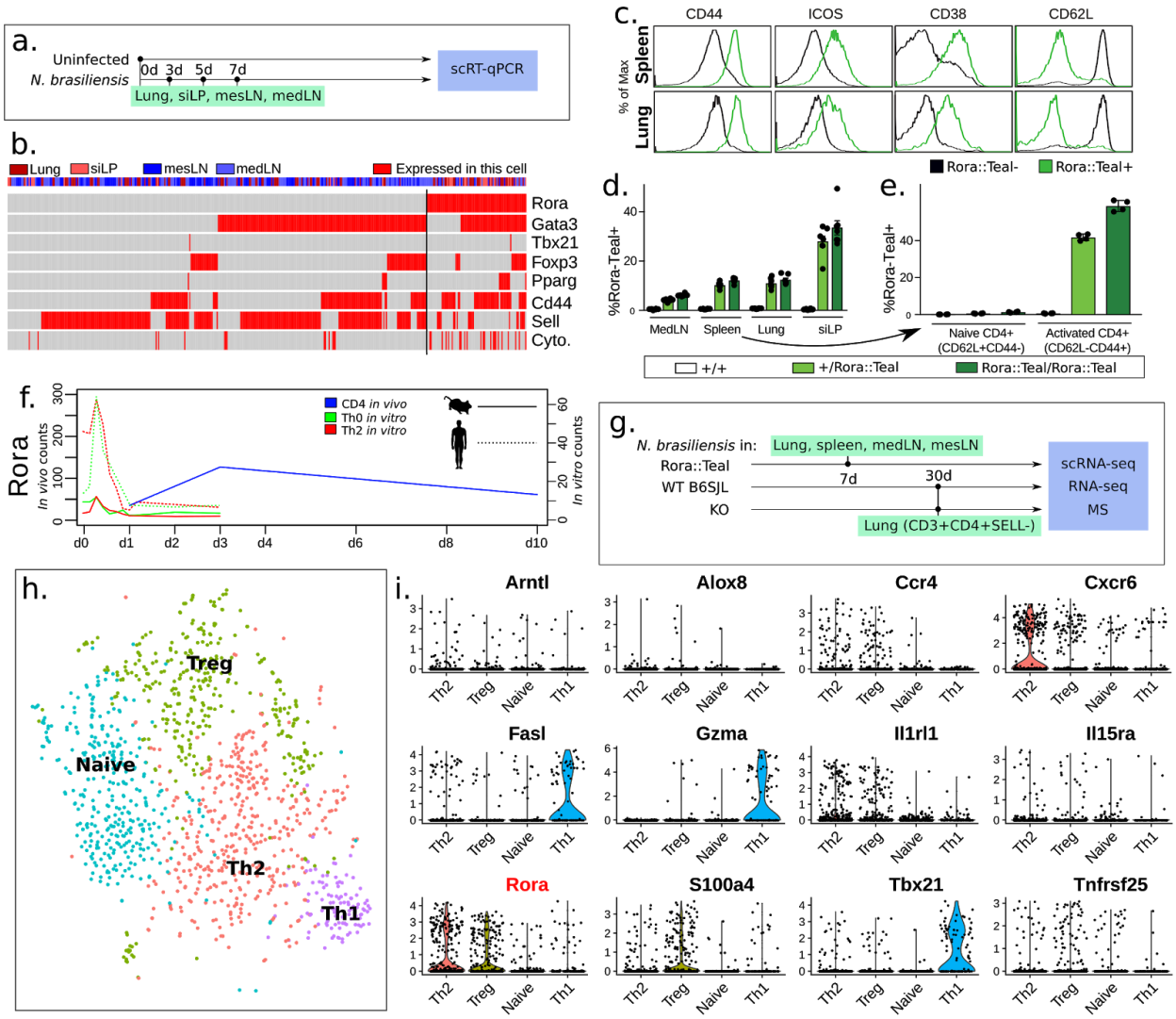

Fig. 2: *Rora* is expressed in activated T cells *in vivo*, with mainly spatial variation. (a) Experimental design for single-cell RT-qPCR. **(b)** Gene expression according to RT-qPCR from CD4 T cells during *N. brasiliensis* infection. *Gata3*^{hi} cells, which we assume are Th2, are also expressing *Rora*. The category "Cytokines" mean that any of *Ifng, Il4, Il13, Il5, Il6, Il10* or *Il17a* are expressed. Top row denotes spatial location: Lung, gut (siLIP), mesenteric and mediastinal lymph nodes. **(c)** Flow cytometric analysis of CD4+ T cells from spleen and lung of Rora$^{+/teal}$ mice (black histogram = Rora-Teal– cells, green histogram = Rora-Teal+ cells). **(d)** Frequency of Rora-Teal expression across tissues, with siLIP having the highest number. **(e)** Fraction of naive or activated cells expressing Rora-Teal. RORA is mainly expressed in activated cells *in vivo*. **(f)** *N. brasiliensis Rora* expression over time, overlaid with previously published mouse and human *in vitro* time course data of *Rora*[7], showing no or low expression of *Rora* in activated *in vitro* T cells. This suggests that a Rora-inducing element is missing in the *in vitro* system. **(g)** Experimental design of scRNA-seq experiment, with **(h)** tSNE clustering of CD4 cells extracted from *N. brasiliensis* infected mice and measured by single-cell RNA-seq. We primarily identify Naive, Th1, Th2 and Treg cells. **(i)** Expression level of some genes, including *Rora*, across different clusters.

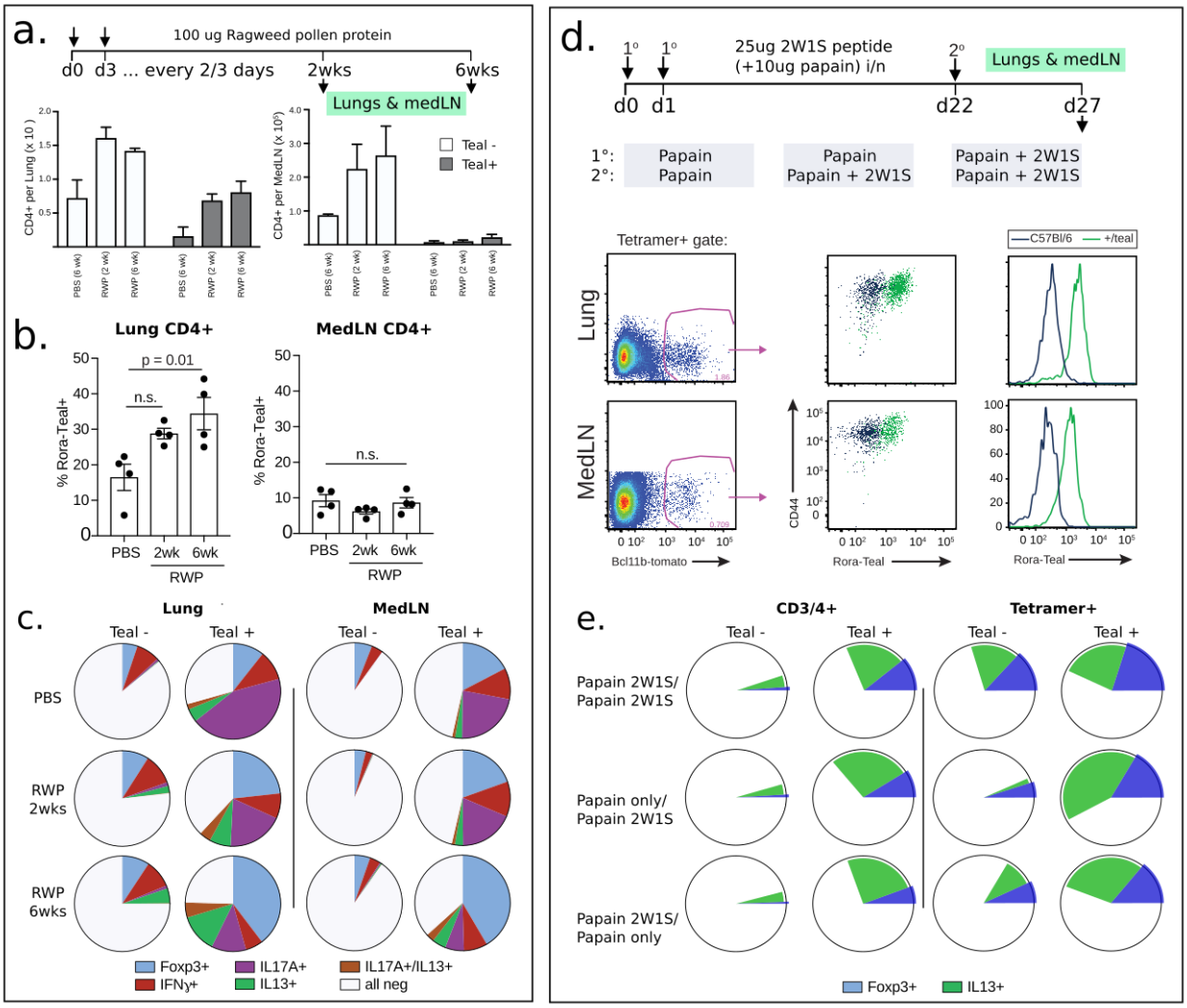

Fig. 3: *Rora* is generally expressed in Th2 during type 2 immunity (a) Flow cytometry analysis of CD4 T cells from mice challenged with Ragweed pollen protein for 2 or 6 weeks, showing that *Rora* is expressed also during allergy. **(b)** RORA expression in CD4 cells increase in lung but not MedLN during the ragweed challenge. **(c)** Flow cytometry analysis diagrams showing higher proportions of cytokines-expression and Tregs cells in Rora+ (Teal+) sorted cells *vs* the Rora- (Teal-), after Ragweed pollen protein exposure. **(d)** Flow cytometry analysis when T cells are challenged with 2W1S/papain, again confirming the presence of *Rora* in tetramer+ cells. Note that we only performed this experiment once. **(e)** Rora+ cells express more FOXP3 and IL13, as compared to RORA- cells, after 21WS/Papain application, showing the presence of Treg and activated cells within the RORA+ population.

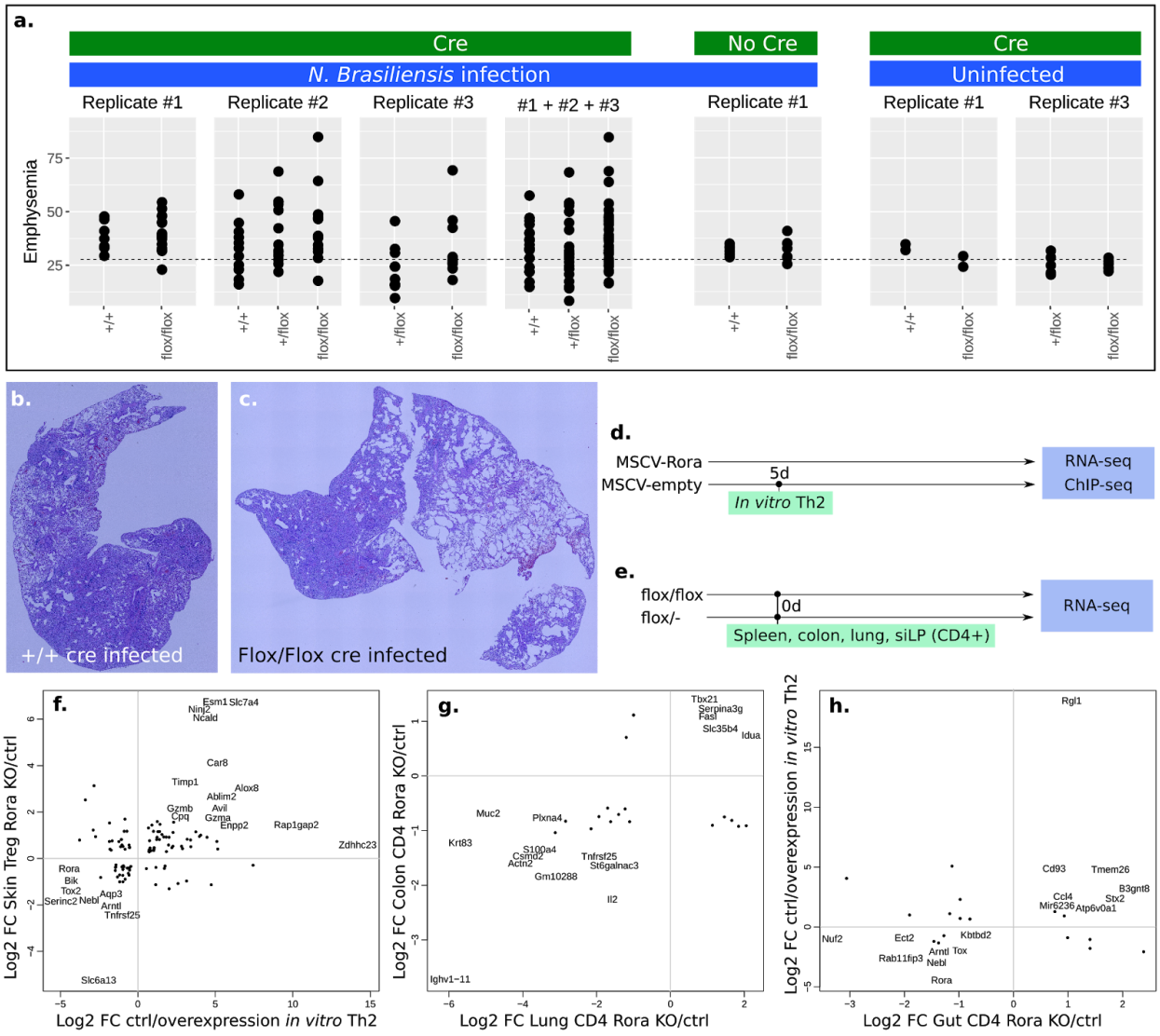

Fig. 4: ***Rora*** **KO affects lung inflammation severity (a)** Assessment of lung emphysema after *N.brasiliensis* infection by microscopy and automated image analysis. Controls are included for no-Cre (against the Rora-flox) and no infection. **(b-c)** Least and most inflamed lung sections as detected by the algorithm, located in +/+ and flox/flox (control and KO). **(d)** Design of the *in vitro* overexpression experiment. **(e)** Design of the *in vivo* KO experiment. **(f)** Overlap of DE genes in *Rora* overexpression *in vitro* Th2 cells *vs* DE genes in Tregs from skin. To increase readability, only some gene names are written out. **(g)** Overlap of DE genes between CD4+ T cells from lung *vs* DE genes from CD4+ T cells from colon (activated T cells). **(h)** Overlap of DE genes between CD4+ T cells from siLP (activated T cells) *vs* DE genes in *Rora* overexpression *in vitro* Th2 cells.

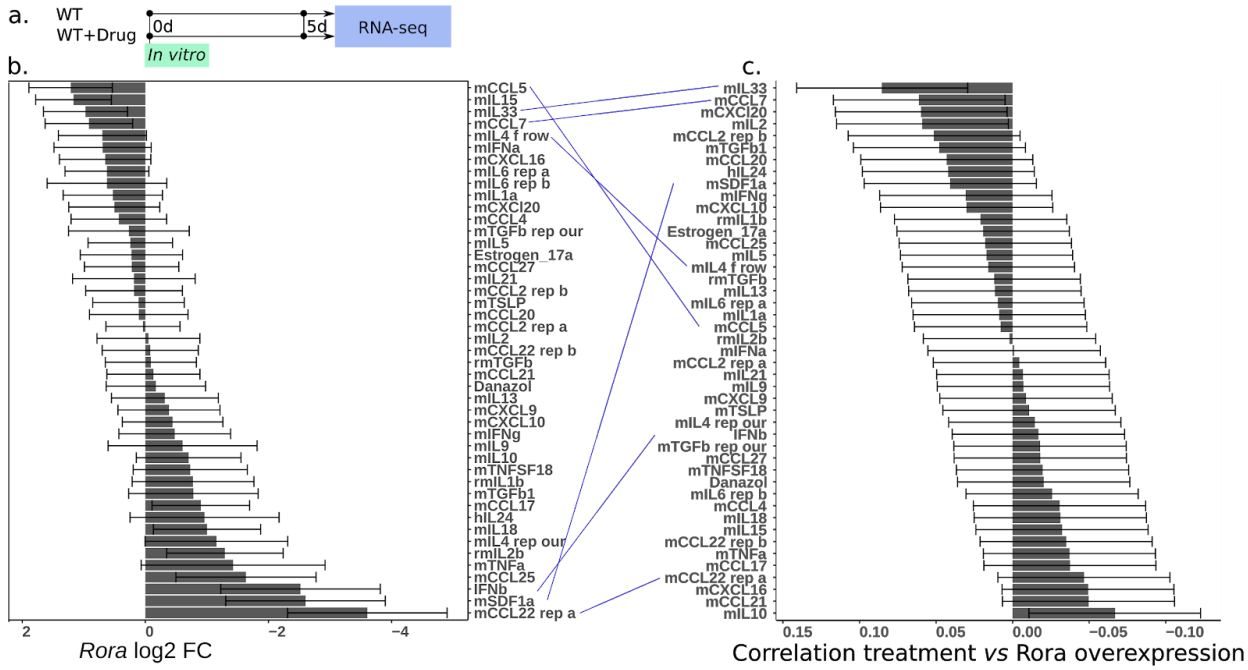

Fig. 5: A screen of cytokine-treated T helper cells reveals genes upstream of *Rora*. The full list of treatments, and the effect on any gene of interest, is provided in Supplementary File S2. **(a)** Design of the *in vitro* experiment **(b)** Effect of all tested conditions on *Rora* expression. **(c)** Global correlation of gene fold change in each treatment, with fold change of gene expression during *Rora* overexpression. Lines have been drawn across panels b-c to highlight how similar a treatment is to directly perturbing *Rora*. Corresponding conditions are thus suggested to act through *Rora* as an early intermediate.

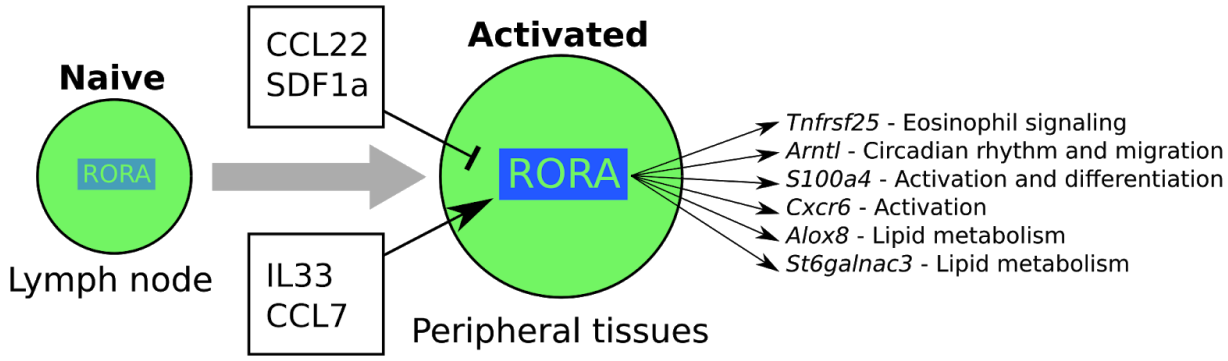

Fig. 6: Summary model. *Rora* is expressed upon activation, and relies on additional stimuli from peripheral tissues for full induction of gene expression. Several cytokines have an impact on *Rora* expression level. *Rora* affects a diverse set of genes of importance for activated T cells, including *Tnfrsf25* (negative regulation of Eosinophils during infection), *Arntl* (circadian rhythm), *S100a4* (activation), *Alox8* and *Stgalnac3* (lipid metabolism).

References

1. Ouyang, W., Kolls, J. K. & Zheng, Y. The biological functions of T helper 17 cell effector cytokines in inflammation. *Immunity* **28**, 454–467 (2008).
2. Halim, T. Y. F. *et al.* Retinoic-acid-receptor-related orphan nuclear receptor alpha is required for natural helper cell development and allergic inflammation. *Immunity* **37**, 463–474 (2012).
3. Wong, S. H. *et al.* Transcription factor RORα is critical for nuocyte development. *Nat. Immunol.* **13**, 229–236 (2012).
4. Lo, B. C. *et al.* The orphan nuclear receptor RORα and group 3 innate lymphoid cells drive fibrosis in a mouse model of Crohn's disease. *Sci Immunol* **1**, eaaf8864 (2016).
5. Han, Y.-H. *et al.* RORα Induces KLF4-Mediated M2 Polarization in the Liver Macrophages that Protect against Nonalcoholic Steatohepatitis. *Cell Rep.* **20**, 124–135 (2017).
6. Malhotra, N. *et al.* RORα-expressing T regulatory cells restrain allergic skin inflammation. *Sci Immunol* **3**, (2018).
7. Henriksson, J., Chen, X., Gomes, T., Ullah, U. & Meyer, K. B. Genome-wide CRISPR screens in T helper cells reveal pervasive cross-talk between activation and differentiation. *BioRxiv* (2017).
8. Stubbington, M. J. T. *et al.* An atlas of mouse CD4+ T cell transcriptomes. *Biol. Direct* **10**, 14 (2015).
9. Lönnberg, T. *et al.* Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves Th1/Tfh fate bifurcation in malaria. *Sci Immunol* **2**, (2017).
10. Miragaia, R. J., Gomes, T., Chomka, A., Jardine, L. & Riedel, A. Single cell transcriptomics of regulatory T cells reveals trajectories of tissue adaptation. *bioRxiv* (2017).
11. Nemeth, K. *et al.* Bone marrow stromal cells use TGF-β to suppress allergic responses in a mouse model of ragweed-induced asthma. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5652–5657 (2010).
12. Moon, J. J. *et al.* Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* **27**, 203–213 (2007).
13. Fang, L., Adkins, B., Deyev, V. & Podack, E. R. Essential role of TNF receptor superfamily 25 (TNFRSF25)

in the development of allergic lung inflammation. *J. Exp. Med.* **205**, 1037–1048 (2008).

14. Wang, J. *et al.* NINJ2- A novel regulator of endothelial inflammation and activation. *Cell. Signal.* **35**, 231–241 (2017).
15. Castro, G. *et al.* RORγt and RORα signature genes in human Th17 cells. *PLoS One* **12**, e0181868 (2017).
16. Bollinger, T. *et al.* Circadian clocks in mouse and human CD4+ T cells. *PLoS One* **6**, e29801 (2011).
17. Akashi, M. & Takumi, T. The orphan nuclear receptor RORα regulates circadian transcription of the mammalian core-clock Bmal1. *Nat. Struct. Mol. Biol.* **12**, 441 (2005).
18. Mandai, Y. *et al.* S1711 Distinct Roles for Cxcr6 and Cxcr6− CD4 T Cells in the Pathogenesis of Chronic Colitis. *Gastroenterology* **138**, S–258 (2010).
19. Latta, M., Mohan, K. & Issekutz, T. B. CXCR6 is expressed on T cells in both T helper type 1 (Th1) inflammation and allergen-induced Th2 lung inflammation but is only a weak mediator of chemotaxis. *Immunology* **121**, 555–564 (2007).
20. Bitsch, F. *et al.* Identification of natural ligands of retinoic acid receptor-related orphan receptor alpha ligand-binding domain expressed in Sf9 cells--a mass spectrometry approach. *Anal. Biochem.* **323**, 139–149 (2003).
21. Harnett, W., Harnett, M. M. & Byron, O. Structural/functional aspects of ES-62--a secreted immunomodulatory phosphorylcholine-containing filarial nematode glycoprotein. *Curr. Protein Pept. Sci.* **4**, 59–71 (2003).
22. Tateyama, M., Oyaizu, N., McCloskey, T. W., Than, S. & Pahwa, S. CD4 T lymphocytes are primed to express Fas ligand by CD4 cross-linking and to contribute to CD8 T-cell apoptosis via Fas/FasL death signaling pathway: Presented in part at the 12th World AIDS Conference, Geneva, Switzerland, 1998. *Blood* **96**, 195–202 (2000).
23. Velaga, S. *et al.* Granzyme A Is Required for Regulatory T-Cell Mediated Prevention of Gastrointestinal Graft-versus-Host Disease. *PLoS One* **10**, e0124927 (2015).
24. Weaver, C. T., Hatton, R. D., Mangan, P. R. & Harrington, L. E. IL-17 family cytokines and the expanding diversity of effector T cell lineages. *Annu. Rev. Immunol.* **25**, 821–852 (2007).
25. Yasuhara, R. *et al.* The β-catenin signaling pathway induces aggressive potential in breast cancer by

up-regulating the chemokine CCL5. *Exp. Cell Res.* **338**, 22–31 (2015).

26. Khaled, A. R. & Durum, S. K. Lymphocide: cytokines and the control of lymphoid homeostasis. *Nat. Rev. Immunol.* **2**, 817–830 (2002).
27. Lodolce, J. P. *et al.* IL-15 Receptor Maintains Lymphoid Homeostasis by Supporting Lymphocyte Homing and Proliferation. *Immunity* **9**, 669–676 (1998).
28. Hunter, C. A. & Jones, S. A. IL-6 as a keystone cytokine in health and disease. *Nat. Immunol.* **16**, 448–457 (2015).
29. Peine, M., Marek, R. M. & Löhning, M. IL-33 in T Cell Differentiation, Function, and Immune Homeostasis. *Trends Immunol.* **37**, 321–333 (2016).
30. Qin, X.-J. *et al.* CCL22 recruits CD4-positive CD25-positive regulatory T cells into malignant pleural effusion. *Clin. Cancer Res.* **15**, 2231–2237 (2009).
31. Vulcano, M. *et al.* Dendritic cells as a major source of macrophage-derived chemokine/CCL22 in vitro and in vivo. *Eur. J. Immunol.* **31**, 812–822 (2001).
32. Campbell, J. J., O'Connell, D. J. & Wurbel, M.-A. Cutting Edge: Chemokine receptor CCR4 is necessary for antigen-driven cutaneous accumulation of CD4 T cells under physiological conditions. *J. Immunol.* **178**, 3358–3362 (2007).
33. Arojo, O. A. *et al.* Active mTORC2 Signaling in Naive T Cells Suppresses Bone Marrow Homing by Inhibiting CXCR4 Expression. *J. Immunol.* (2018). doi:10.4049/jimmunol.1800529
34. He, W. *et al.* Circadian Expression of Migratory Factors Establishes Lineage-Specific Signatures that Guide the Homing of Leukocyte Subsets to Tissues. *Immunity* (2018). doi:10.1016/j.immuni.2018.10.007
35. Druzd, D. *et al.* Lymphocyte Circadian Clocks Control Lymph Node Trafficking and Adaptive Immune Responses. *Immunity* **46**, 120–132 (2017).
36. Kim, K. *et al.* RORα controls hepatic lipid homeostasis via negative regulation of PPARγ transcriptional network. *Nat. Commun.* **8**, 162 (2017).
37. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
38. Weigmann, B. *et al.* Isolation and subsequent analysis of murine lamina propria mononuclear cells from

colonic tissue. *Nat. Protoc.* **2**, 2307–2311 (2007).

39. Tucker, M. S., Karunaratne, L. B., Lewis, F. A., Freitas, T. C. & Liang, Y.-S. Schistosomiasis. *Curr. Protoc. Immunol.* **Chapter 19**, Unit 19.1 (2001).
40. Mann, V. H., Morales, M. E., Rinaldi, G. & Brindley, P. J. Culture for genetic manipulation of developmental stages of Schistosoma mansoni. *Parasitology* **137**, 451–462 (2010).
41. Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics* **26**, 979–981 (2010).
42. Benaglia, T., Chauveau, D., Hunter, D. & Young, D. mixtools: An R package for analyzing finite mixture models. *J. Stat. Softw.* **32**, 1–29 (2009).
43. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
44. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* (2017). doi:10.1038/nmeth.4197
45. Butler, A. & Satija, R. Integrated analysis of single cell transcriptomic data across conditions, technologies, and species. *bioRxiv* (2017).
46. Stubbington, M. J. T. *et al.* T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **13**, 329–332 (2016).
47. Bürglin, T. R. & Henriksson, J. FACSanadu: Graphical user interface for rapid visualization and quantification of flow cytometry data. *bioRxiv* 201897 (2017). doi:10.1101/201897
48. Folch, J., Lees, M. & Sloane Stanley, G. H. A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* **226**, 497–509 (1957).
49. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

Methods

Mice strains

Rora reporter mice

Rora T2A Teal (manuscript in preparation; JW and ANJM) were generated briefly as follows: A reporter cassette, encoding a short Gly-Ser-Gly linker peptide, FLAG epitope tag, T2A self-cleaving peptide and Teal fluorescent protein, followed by a loxP-flanked Neomycin cassette, was inserted (directly upstream of the) *Rora* stop codon. Successful targeting of JM8 ES cells was confirmed by Southern blot and the Neomycin cassette was removed from the resultant mice by inter-crossing with a Cre recombinase strain.

Rora KO mice

STOCK Tg(Cd4-cre)1Cwi/BfluJ mice were crossed with $Rora^{tm1a(EUCOMM)Wtst}$ mice. F1 progenies were crossed again, and F2 were tested for having at least one allele of Cd4-cre, and two alleles of *Rora-lox*.

Ethics statement

The mice were maintained under specific pathogen-free conditions at the Wellcome Trust Genome Campus Research Support Facility or the Medical Research Council (MRC) Ares facility (Cambridge, UK). These animal facilities are approved by and registered with the UK Home Office. All procedures were in accordance with the Animals (Scientific Procedures) Act 1986. The protocols were approved by the Animal Welfare and Ethical Review Body of the Wellcome Trust Genome Campus.

*O*verexpression of *Rora* and analysis

Rora was cloned into a retroviral vector (M6-mCherry-Cter) as follows. cDNA was made from RNA extracted from mouse spleens, following the protocol of SuperScript II reverse transcriptase (Thermo Fisher Scientific #18064014). Primers were designed against *Rora* isoform ENSMUST00000113624 (fwd: ATGTATTTTGTGATCGCAGCGATGAAAGCTCAAATTGAA, rev: TTACCCATCGATTTGCATGGCTGGCTCAAATT). After PCR and gel purification, another round of PCR was done to add compatible overhangs (fwd: CGATaagcttATGTATTTTGTGATCGCAGC, rev: GaatgcggccgcCCCATCGATTTGCATGG). The product and backbone were digested with HindIII-HF and NotI-HF (NEB), purified, and ligated using T4 ligase (NEB). Several clones were picked and sequence verified by sanger sequencing (M6-Rora2-mCherry). Two other constructs were also constructed analogously for ENSMUST00000034766 (M6-Rora1-mCherry and M6-Rora1-GFP).

The virus production, activation and transfection was done as previously, with addition of IL4[7]. The transfection was done into 3 biological replicates (JaxJ, 8 weeks old), each with the 3 constructs and the no-insert M6 as control. No puromycin was used for selection. Cells were harvested on day 4, FACS sorted, and libraries prepared using the Sanger institute standard pipeline. Sequencing was done on a HiSeq 2500.

The differentially expressed genes were called by DESeq2[37] (normalized counts ~ treatment + mouse). Each of the constructs were analyzed. Of these, one had *Rora* as the most DE gene and was retained for further comparison (M6-Rora2-mCherry). Counts and condition matrices are provided in Supplementary File S2.

Mice infection and cell isolation

Nippostrongylus Brasiliensis infection

C57BL/6 female mice were subcutaneously injected with 100ul (300 live third stage *N. Brasiliensis* larvae per dose), over two sites. MedLNs, mesLNs, spleens, lungs, or siLP were taken from infected mice, 3, 5, 7, 10 or 30 days after infection as well as from uninfected mice. At each time point, cells were isolated from medLNs and mesLNs, by smashing the tissue through 70µm cell strainer. Lungs were incubated in collagenase D (0.72mg/ml, Amersham, Bucks, UK) for 30min, smashed through 70µm cell strainer, and suspended in RBC lysis buffer (eBioscience Ltd). Spleens were smashed through 70µm cell strainer, and suspended in RBC lysis buffer. siLP cells were isolated from colonic tissue as previously described[38]. In brief, siLP cells were released by digestion of the tissue with RPMI/4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) supplemented with 60 µg/ml DNaseI (Sigma), and 400 ng/ml of Liberase (Roche Applied Science, Burgess Hill, UK). Isolated cells were stained with conjugated antibodies for sorting (Supplementary File S3). For RT-qPCR and bulk RNA-seq experiments, single cells were sorted into lysis buffer and was stored in -80°C.

Schistosoma mansoni infection

The complete life cycle of the parasite *S. mansoni* is maintained at the Wellcome Sanger Institute. Cercariae, the mammalian infective stage, were harvested by exposure of infected *Biomphalaria glabrata* snails to light for two hours in aquarium-conditioned water. Female C57BL/6 mice were infected with 300-mix sex *S. mansoni* cercariae via IP as described elsewhere[39]. Mice were checked twice a day for any sign of health deterioration. After 6 weeks post infection, mice were culled with an overdose of anaesthetic, adult worms were perfused from the mouse circulatory system as described previously[40], and mouse spleen and mesLNs were recovered surgically and placed in PBS for further processing. Cells were isolated from tissues and were stained for sorting. Activated T cells, CD3+CD4+CD44+CD62L-, were sorted into a C1 chip and processed for RNA-seq.

Lung immune challenge models

Mice were anaesthetised by isofluorane inhalation followed by intranasal administration of Ragweed pollen (100 µg per dose; Greer Laboratories, Lenoir, NC), papain (10 µg per dose; Sigma) or 2W1S peptide (10 µg per dose; Designer Bioscience), dissolved in 40 µl PBS at the indicated time points. The PE-conjugated I-A(b) 2W1S tetramer was obtained from the NIH Tetramer core.

Cell sorting

Cells were sorted with a BD Influx Cell Sorter or SY3200 Synergy cell sorter (iCyt) and analysed on Fortessa or LSRII BD cell analyzers using FlowJo. Surface and intracellular stainings were carried out according to the eBioscience protocols. A full list of the antibodies used is provided in Supplementary S4.

Histology

Lung lobes were fixed for 48 hours in 10% neutral-buffered formalin (Sigma), washed in PBS and transported to AML Laboratories (Florida) for embedding, sectioning and staining with haematoxylin and eosin.

Microscopy and image analysis

Images were captured using a VHX keyence microscope, using objective #2D1510064 at magnification "200". The 2D stitching feature was used to capture the entirety of each lung section. Two sections were imaged from each mouse, with focus automatically calibrated for each section. All images were acquired in one seating, with the same settings for brightness.

Image analysis was done in R using the EBimage package[41]. In brief, pixels corresponding to background, cytoplasm and nuclei were manually annotated in one image. A classifier was set up to separate pixels into the 3 classes based on manual inspection of the RGB space.

To estimate emphysema, the distribution of inner holes was calculated by first detecting non-background area. A binary dilation operation was applied to filter out 1-2 pixel holes caused by improper tissue detection. Segmentation was then performed to find disjunct regions. The area of each region was calculated.

From the histogram H of hole sizes, we estimate the unevenness (indicating emphysema, compensating for overall compression of slides due to fixation) as $S=H_{99\%}/H_{80\%}$. Other ranges where tested and give similar outcome. Veins, surrounding space, and other artifacts are filtered out above the 1% top percentile, while the lower percentile reflects tissue shrinkage. Alternatively, the score can be interpreted as a measure of hole size unevenness/variance. To test for emphysema in the *Rora* KO we fit a linear model S ~ replicate + genotype, with genotype=1 for Cre and +/Flox, 2 for Cre and Flox/Flox, and 0 otherwise (p=0.018). A simpler t-test of +/+ vs Flox/Flox, not using any of the +/Flox samples, and ignoring replicate batch effects, gives p=0.17. This is not significant but in the direction of the more inclusive linear model.

In vivo bulk RNA-seq

RNA was extracted with SPRI beads, followed by Smartseq-2. Read counts were estimated with Kallisto. DE genes were estimated by DESeq2[37] after counts had been rounded to the nearest integer count. Simple control *vs* treatment linear models were used. Counts and condition matrices are provided in Supplementary File S2.

Quantitative single-cell gene expression analysis

Single cell RT-qPCR measurement and analysis

Single-cell gene expression analysis was performed using BioMark 96.96 Dynamic Array platform (Fluidigm, San Francisco, CA) and TaqMan Gene Expression Assays (Applied Biosystems, Carlsbad, CA). Single cells were sorted into 5µl of CellsDirect reaction mix and immediately stored in -80C. Control wells containing no cells were included. On thawing, a mix containing 2.5µL gene specific 0.2x

TaqMan gene expression assays (Applied Biosystems), 1.2 µL CellsDirect RT/Taq mix, and 0.3 µL TE buffer were added to each well. RT–PCR pre-amplification cycling conditions were: 50°C, 15min; 95°C, 2min; 22 x(95°C, 15s; 60°C, 4min). Samples were diluted 1:5 in TE buffer and 6% were mixed with TaqMan Universal PCR Master Mix (Applied Biosystems). The sample mix and TaqMan assays were loaded separately into the wells of 96.96 Gene expression Dynamic Arrays (Fluidigm) in presence of appropriate loading reagents. The arrays were read in a Biomark analysis system (Fluidigm). ΔCt values were calculated in reference to the average of *Atp5a1*, *Hprt1* and *Ubc*.

The expression of each gene was fit by a bivariate normal distribution (R mixtools package[42]) and a cut-off set at the average position between the lower and upper gaussian midpoints. Genes above this were considered expressed iff above this cut-off. qPCR levels and metadata are provided in Supplementary File S1.

S. mansoni C1 single cell RNA-seq and analysis

For *S. mansoni*, three small (5–10 µm) C1 Single-Cell Auto Prep IFC chips (Fluidigm) were primed and 5000 cells were sorted directly into the chip. To allow estimation of technical variability, 1 µl of a 1:4000 dilution of ERCC (External RNA Controls Consortium) spike-in mix (Ambion, Life Technologies) was added to the lysis reagent. Cell capture sites were visually inspected one by one using a microscope. The capture sites that did not contain single cells were noted and were removed from downstream analysis. Reverse transcription and cDNA preamplification were performed using the SMARTer Ultra Low RNA kit (Clontech) and the Advantage 2 PCR kit according to the manufacturer's instructions on the C1 device. cDNA was harvested and diluted to 0.1–0.3 ng/µl and libraries were prepared in 96-well plates using a Nextera XT DNA Sample Preparation kit (Illumina) according to the protocol supplied by Fluidigm. Libraries were pooled and sequenced on an Illumina HiSeq2500 using paired-end 75-bp reads.

Salmon was used to estimate gene expression counts. Poor quality libraries were eliminated using Scater based on exonic and mitochondrial read counts. For all queries, a gene was considered expressed if the normalized count was above 0.5. Gene overlap was tested using Fisher's method. Counts and condition matrices are provided in Supplementary File S3.

N. brasiliensis C1 single cell RNA-seq

T cells from *N. brasiliensis* infected mice were captured with C1 analogously to the *S. mansoni* mice. The cells were sorted for CD3+CD4+SELL-, from mesLN, medLN, lung and spleens. This data is available but was not used due to low quality.

N. brasiliensis Smartseq-2 single-cell RNA-seq and analysis

For this analysis we included CD3+CD4+SELL- cells from the lungs at 30 days after infection, CD3+CD4+Rora-TEAL+ and CD3+CD4+Rora-TEAL- cells from *N. brasiliensis* infected *Rora*$^{+/teal}$l reporter mice. Further we include cells from lungs, spleens, medLN and mesLN from infected and uninfected *Rora*$^{+/teal}$ mice, 7 days after infection.

Single cell transcriptomes were generated by using the Smartseq-2 protocol[43]. Single cells were sorted into 96 well plate that contained 5µl of of Triton-X lysis buffer as described in Picelli et al., 1 µl of 10 µM olido-dT30-VN, 1 µl dNTP mix (25 mM each) and ERCC controls at a final dilution of 1:64m and

immediately stored in -80C. cDNA was submitted for the amplification for 25 cycles on Alpha Cycler 4 thermal cycler. Amplified cDNA went through two subsequent rounds of cleaning by using Agencourt AMPure XP beads (Beckman Coulter UK Ltd, High Wycombe, UK) at a 1.0x ratio on a liquid handler Zephyr G3 NGS Workstation (PerkinElmer) and eluted in 20ul of RNAse free water. Purified cDNA was subjected to quality control using 1 µl of cDNA on an Agilent 2100 BioAnalyser (Agilent Technologies, Santa Clara, CA, USA) using the Agilent High Sensitivity DNA kit. Samples were normalised to a concentration of 0.3 ng/µl. Nextera libraries were prepared using Nextera XT DNA Sample Preparation kit (Illumina) according to the protocol supplied by Fluidigm. The library preparation has been done in combination with benchtop liquid handlers: Zephyr G3 NGS Workstation (PerkinElmer) and Mantis (Formulatrix). Purified pool of Nextera libraries was subjected to quality control using 1 µl of cDNA on an Agilent 2100 BioAnalyser using the Agilent High Sensitivity DNA kit.

Salmon v0.8.2[44] was used to map the reads to ensembl cDNA reference GRCm38 to quantify the gene expression. Counts were used for all the analyses. Low quality cells with >15% mitochondrial genes or <1500 total genes were filtered out. Seurat[45] was used for single-cell analysis: Highly variable genes were calculated (x.low.cutoff=0, x.high.cutoff=5, y.cutoff=0.1), PCA performed on these (pcs.compute=50), and t-SNE on the first 25 PCA components. Finally clustering was done (reduction.type="pca", dims.use=1:25, resolution=0.5). Common marker genes were plotted on top of the clusters and the clusters identities were assigned manually (*Gata3*, *Tbx21*, *Foxp3*, *Sell,* and *Rorc*). The TCR sequences for each single T cell were assembled using TraCeR[46] which allowed the reconstruction of the TCRs from scRNA-seq data and their expression abundance, in addition to identification of the size, diversity and lineage relation of clonal subpopulations. Cells for which more than two alpha or beta chains were identified were excluded from further analysis. iNKT cells were detected by their characteristic TCRA gene segments (TRAV11–TRAJ18). Only CD4+ T cells were kept in the final plot; other cells remain annotated in the ArrayExpress submission. Counts and condition matrices are provided in Supplementary File S2.

Screening for upstream regulators of *Rora*

Cell culture

Naive CD4+ T helper cells were extracted from spleens of JaxJ mice, 8 weeks old males (Stem Cell Technologies #19765), according to the manufacturer's protocol. Up to 4 spleens were pooled in each biological replicate. The cells were activated in 96-well round bottom plates that were coated with 3ul/ml anti-CD3e (BioLegend #100202) and 5ul/ml anti-CD28 (BioLegend #102102) for 4 hours at 32°C. 24 hours later cytokines were added (See Supplementary file S4). 72 hours after activation the cells were pelleted, resuspended in RLT buffer (Qiagen #79216), and stored at -80°C.

For the repeated screen, culturing was done in the same manner. The cells were then stained with PI and anti-CD4 eFluor 660 (eBioscience 50-0041-82, clone GK1.5) for 30 minutes in IMDM media. The cells were then FACS sorted to retain live CD4+ T cells, on average 5000 cells per well.

RNA-sequencing

For the first screen, RNA was extracted using Agencourt AMPure XP beads (Catalog #A63881) at 1.5:1 ratio into 10ul elution volume. RNA-seq was done following Smart-seq2[43] but with the following changes: The input was increased to 7ul and Mix 1 consisted of 2ul 10uM dNTP and 2.5ul 10uM oligo-dT. Mix 2 consisted of 5ul SmartScribe buffer, 5ul 5M Betaine, 1.25ul SmartScribe, 1.25ul (100uM) DTT, 0.15ul 1M MgCl2, 0.38ul 100uM TSO, and 0.47ul SUPERase inhibitor. SPRI cleanup was done at ratio 0.8:1 following first-strand synthesis, eluted into 12ul NFW. Mix 3 consisted of 12.5ul KAPA HiFi 2x master mix and 0.5ul 10uM ISPCR. PCR was done with 10 cycles. In the second screen, RNA-seq was done similarly with the following modification: No SPRI-cleanup was done after first-strand synthesis. Instead a larger amount of mix 3 was added directly.

On average 0.2ng cDNA was then used for Nextera XT library preparation, done at 25% of volume recommended by manufacturers protocol. Sequencing was done with 2 lanes of Illumina 50bp PE HiSeq 2500.

Upstream analysis

Salmon was used to estimate gene expression counts. Poor quality libraries were eliminated using Scater based on exonic and mitochondrial read counts. Genes with average normalized counts below 0.3 were removed. The increase/decrease in *Rora* expression level was computed by DEseq2[37] using a linear model: normalized counts ~ treatment + mouseReplicate, where treatment is a factorial variable. The correlation heatmap between treatments was calculated as cor(cor(FC)). The first correlation is used as a dimensionality reduction while the second correlation is used to remove the technical bias toward positive correlation. Clustering was done using R hclust, standard parameters.

The CellTrace FACS data was analyzed with FACSanadu[47] to obtain cells in each division stage. The cell counts were analyzed across replicates and compared to controls using R.

Metabolite LC-MS

Metabolites were extracted using a modification of a previous method[48]. Briefly, roughly 200k cells were resuspended in chloroform/methanol (2:1). After mixing samples, water was added (300 µL) and centrifuged at 13,000 x *g* for 20 min. The organic layer (lower fraction) was collected and dried under a stream of nitrogen. The organic fraction was reconstituted in chloroform/methanol (1:1; 100 µL) and 10 µL were mixed in 90 µL IPA/acetonitrile/water (2:1:1). The samples were analysed by LC-MS using an LTQ Orbitrap Elite Mass Spectrometer (Thermo Scientific). Sample (5 µL) was injected onto a C18 packed-tip column (75µm x 100mm, Thermo Scientific) at 55°C. The mobile phase A was acetonitrile/water 60:40, 10 mM ammonium formate whilst B was LC–MS-grade acetonitrile/isopropanol 10:90, 10 mmol/L ammonium formate. In negative ion mode, ammonium acetate was used to aid ionisation. In both positive and negative ion mode, the gradient used (Supplementary File S4) was run at a flow rate 0.5 mL/min. Data were acquired with a mass range of 100-2000 m/z. Chromatograms were converted to mzML format and annotated using an in-house R script. The annotated metabolites were size factor normalized to enable comparison. PCA was performed on the *in vitro* and *in vivo* samples separately (Supplementary Fig. 5). Lungs and spleens from *in vivo* samples appear to cluster together,

suggesting no need to correct for tissue differences. *In vitro* samples where compared with Limma[49], with all naive cells compared to all day 6 samples Th0/1/2. For the *in vivo* samples, we used the simplest model of ~1+isko. Bayesian error correction was performed.