Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells

Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells

Chamutal Bornstein, [1]

Shir Nevo, [1]

Amir Giladi, [1]

Noam Kadouri, [1]

Marie Pouzolles, [2]

François Gerbe, [3]

Eyal David, [1]

Alice Machado, [2]

Anna Chuprin, [1]

Beáta Tóth, [4]

Ori Goldberg, [5]

Shalev Itzkovitz, [4]

Naomi Taylor, [2]

Philippe Jay, [3]

Valérie Zimmermann, [2]

Jakub Abramson, [1]✉

Email jakub.abramson@weizmann.ac.il

Ido Amit, [1]✉

Email ido.amit@weizmann.ac.il

[1] Department of Immunology, Weizmann Institute of Science, Rehovot, Israel

[2] Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, Montpellier, France

[3] IGF, CNRS, INSERM, University of Montpellier, Montpellier, France

[4] Department of Cell Biology, Weizmann Institute of Science, Rehovot, Israel

[5] Department of Pediatrics, Schneider Children's Medical Center, Petach Tikva, Israel

Received: 5 September 2017 / Accepted: 6 June 2018

Abstract

T cell development and selection are coordinated in the thymus by a specialized niche of diverse stromal populations[1, 2, 3]. Although much progress has been made over the years in identifying the functions of the different cell types of the thymic stromal compartment, there is no comprehensive characterization of their diversity and heterogeneity. Here we combined massively parallel single-cell RNA-sequencing[4, 5], spatial mapping, chromatin profiling and gene targeting to characterize de novo the entire stromal compartment of the mouse thymus. We identified dozens of cell states, with thymic epithelial cells (TECs) showing the highest degree of heterogeneity. Our analysis highlights four major medullary TEC (mTEC I–IV) populations, with distinct molecular functions, epigenetic landscapes and lineage regulators. Specifically, mTEC IV constitutes a new and highly divergent TEC lineage with molecular characteristics of the gut chemosensory epithelial tuft cells. Mice deficient in *Pou2f3*, a master regulator of tuft cells, have complete and specific depletion of mTEC IV cells, which results in increased levels of thymus-resident type-2 innate lymphoid cells. Overall, our study provides a comprehensive characterization of the thymic stroma and identifies a new tuft-like TEC population, which is critical for shaping the immune niche in the thymus.

AQ1

A comprehensive characterization of the thymic stroma identifies a tuft-cell-like thymic epithelial cell population that is critical for shaping the immune niche in

the thymus.

These authors contributed equally: Chamutal Bornstein, Shir Nevo, Amir Giladi, Noam Kadouri

These authors jointly supervised this work: Jakub Abramson, Ido Amit

Main

The thymus constitutes a specialized lymphoid organ, where immature T lymphocytes are educated to recognize foreign antigens, while tolerating self[1]. The T cell 'educational program' involves two central steps, which occur in two anatomical compartments of the thymus, the cortex and the medulla. Both compartments are characterized by the presence of specialized stromal cells, which provide the desired microenvironment for different checkpoints in T cell development and selection[2, 3]. Cortical thymic epithelial cells (cTEC) coordinate the early stages of T cell development and positive selection of thymocytes[6]. The later steps of T cell development, including negative selection, are primarily carried out by mTECs[1, 2, 3].

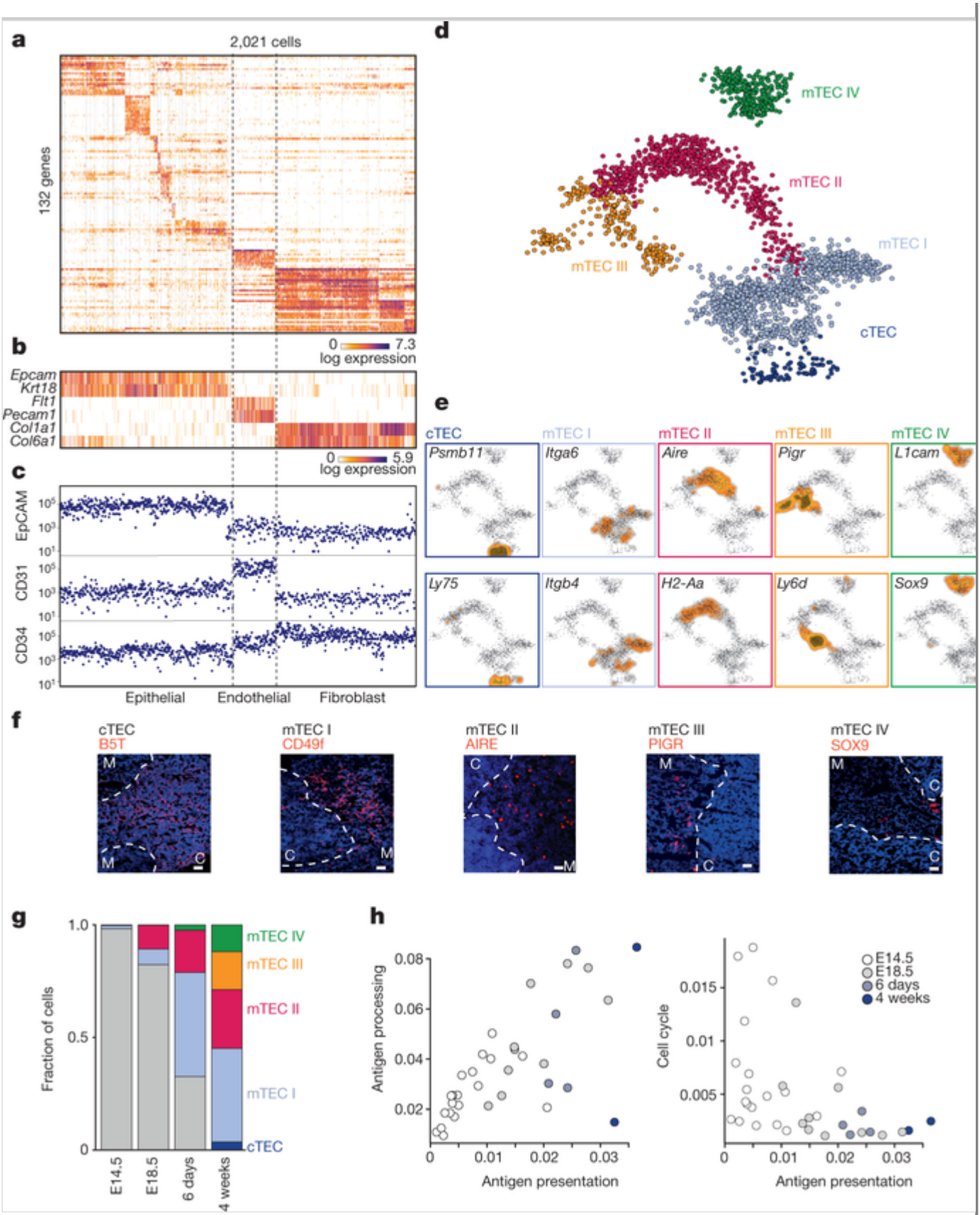
AQ2

Although much progress has been made over the years in elucidating the function of the different cell types of the thymic stroma, their diversity, heterogeneity and molecular pathways are still poorly characterized. To de novo characterize the entire stromal compartment of the thymus, we performed massively parallel single-cell RNA-sequencing (MARS-seq)[4, 5] of 2,021 non-haematopoietic cells ($CD45^{-}$) isolated from adult mouse thymi (Extended Data Fig. 1). In order to link between the canonical surface markers to the single-cell RNA-sequencing data, we used an index sorting strategy that allowed for retrospective analysis of surface markers of each individual cell. We then used the MetaCell pipeline to identify homogeneous and robust groups of cells (Methods). This analysis showed that the thymic stroma is composed of three major lineages, consisting of fibroblasts (*Col1a1* and *Col6a1*), endothelial cells (*Pecam1* and *Flt1*) and epithelial cells (*Epcam* and various keratin genes; Fig. 1a–c and Extended Data Fig. 2). Of the three linages, the epithelial cells displayed the largest heterogeneity in gene-expression programs (Extended Data Fig. 2 and Supplementary Tables 1, 2), suggesting that they are more complex and heterogeneous than previously anticipated.

Fig. 1

The medulla epithelial compartment has diverse molecular functions.

a, Clustering analysis of 2,021 thymic stromal ($CD45^-$) cells from seven biological replicates of 4–6-week-old mice. **b**, Expression of selected marker genes. **c**, Index sorting tracks showing protein level intensity. **d**, Two-dimensional graphical representation of 2,341 single $CD45^-EpCAM^+$ cells separated into five TEC subsets. **e**, Kernel density projection of differentially expressed genes onto the two-dimensional graph. **f**, Immunofluorescence images of thymus sections. Medulla (M) and cortex (C) are separated by a dashed line, distinguished by nuclei density. Blue, DAPI. β5t is also known as PSMB11; CD49f is also known as ITGA6. Scale bars, 40 μm. Images are representative of three independent animals with similar results. **g**, Distribution of the TEC subsets along four developmental time points. Grey labels represent cells distinct from adult TECs. **h**, Progression of early cTECs towards adult cTECs. Circles represent cTEC metacells coloured by the developmental time points of the majority of cells. Axes represent share of each gene module from the entire metacell transcriptome.



In order to comprehensively characterize the TEC compartment, we sorted additional 1,716 $CD45^{-}EpCAM^{+}$ single cells (Extended Data Figs 1, 3). Clustering analysis combined with two-dimensional projection of the epithelial cells from both

datasets revealed dozens of different TEC subpopulations (Fig. 1d, e, Extended Data Fig. 3 and Supplementary Table 2). The TEC subpopulations clustered within five major molecular types, each distributed at a distinct position within the two-dimensional projection (Fig. 1d and Extended Data Fig. 3). Index sorting analysis using the canonical cortical (Ly51) and medullary (UEA1) markers revealed that only one of the groups corresponds to the $Ly51^{+}UEA1^{-}$ population and expresses cTEC-specific genes, including *Prss16*, *Psmb11* and *Ctsl*[2, 3]. By contrast, the other four TEC populations stained positively for UEA1 and had no or low expression of Ly51 (Extended Data Fig. 3), suggesting that these cells reside in the medulla. This was further validated by immunofluorescence staining and single-molecule RNA fluorescence in situ hybridization using a panel of markers specific to the individual TEC subpopulations (Fig. 1f and Extended Data Fig. 3). Therefore, on the basis of these data, we reclassified the mTEC compartment into four major groups (mTEC I–IV), reflecting their distinct transcriptional and molecular characteristics. Specifically, mTEC I is characterized by high expression of *Itga6* and *Sca1* (also known as *Ly6a*) (Fig. 1e and Extended Data Fig. 3), expression of which have previously been associated with putative TEC progenitors[7]. The mTEC II population is characterized by specific expression of the canonical markers of mature mTECs, including high expression of *Aire*, *Fezf2*, *Cd40*, *H2-Aa* or *Cd74* (Fig. 1e and Extended Data Fig. 3). mTEC III represents a heterogeneous population expressing several unique genes (*Pigr*, *Ly6d*, *Spink5*, *Ivl* and *Krt10*), some of which have been linked to a putative population of mTEC that previously expressed AIRE (post-AIRE cells)[8, 9] (Fig. 1e and Extended Data Fig. 3). Notably, the mTEC IV population does not express any classical mTEC or cTEC markers, but rather a unique set of genes such as *Lrmp*, *Avil*, *Trpm5*, *Dclk1*, *Gng13*, *L1cam* and *Sox9* (Fig. 1e and Extended Data Fig. 3).

To investigate the dynamics of the different TEC populations during early development, we performed MARS-seq analysis of 3,074 sorted $CD45^{-}EpCAM^{+}$ single cells isolated from thymi at major developmental stages: embryonic day 14.5 (E14.5), E18.5 and day 6 after birth (Extended Data Figs. 1, 4). Because the developing thymus may have additional cell types or states that are not observed in the adult, we associated embryonic TEC metacells with adult phenotypes only if a large fraction of their cell neighbours was of adult origin (Methods, Fig. 1g, Extended Data Fig. 4 and Supplementary Table 3). This analysis highlighted the dynamic changes in the TEC compartment during thymus organogenesis. Although most of the E14.5 TECs were relatively homogenous and expressed a large number of cTEC-specific genes, their general transcriptional signature was distinct from adult cTECs (Fig. 1g and Extended Data Fig. 4). Specifically, we observed

progressive downregulation of cell cycle genes and upregulation of the MHC-II pathway in the adult compared to the embryonic cTECs (Fig. 1h, Extended Data Fig. 4 and Methods). While none of the adult mTEC subpopulations were present at the E14.5 stage, mTEC I and II, but not mTEC III and IV, became detectable in the thymus at E18.5 (Fig. 1g and Extended Data Fig. 4). At neonatal day 6, mTEC I, II and IV were present, although with different frequencies than in the adult thymus (Fig. 1g and Extended Data Fig. 4). Measuring the percentage of proliferating cells, we found that E14.5 TECs are the most proliferative cells in the thymus, while after birth most of the TEC proliferation is restricted to mTEC I and mTEC II cells (Extended Data Fig. 4).

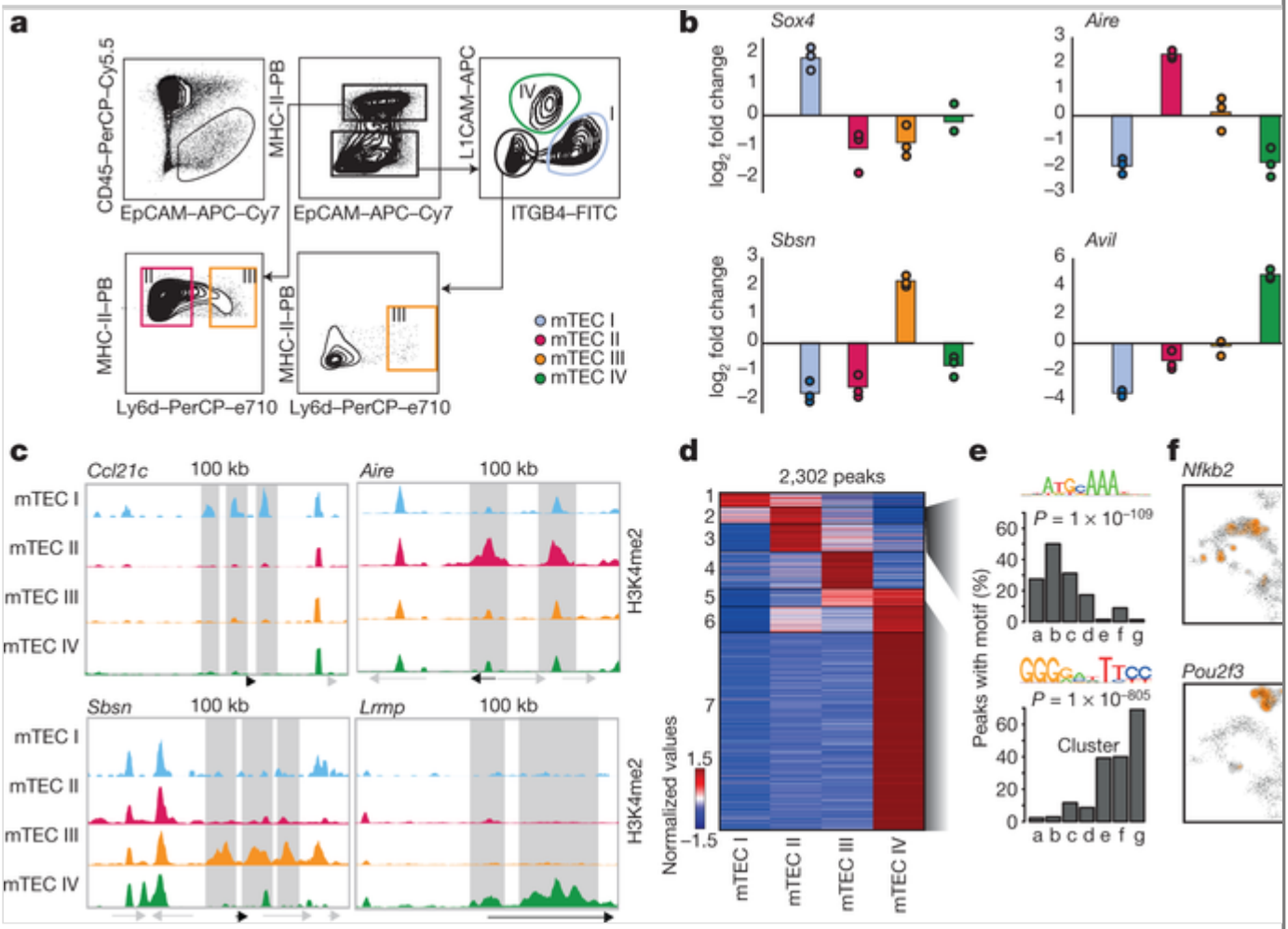
To further characterize the newly defined mTEC subtypes, we established a new sorting strategy, based on a panel of surface markers unique to each population (Fig. 2a). This strategy was validated by MARS-seq and qPCR analyses of sorted $EpCAM^{+}$ populations, which were gated according to this scheme (Fig. 2b and Extended Data Fig. 5). Profiling the putative enhancer regulatory elements (marked by H3K4me2) of the four mTEC populations using indexing-first chromatin immunoprecipitation followed by deep sequencing[10] revealed that each subset is characterized by a unique set of distal enhancer regions with the mTEC IV population showing the most distinct regulatory elements (Fig. 2c, d, Extended Data Fig. 5 and Extended Data Table 1). In order to determine specific transcription-factor-binding sites, we performed an assay of transposase-accessible chromatin and analysed peaks of open chromatin within enhancer regions[11], followed by de novo motif finding (Extended Data Fig. 5 and Extended Data Table 1). This analysis revealed that accessible enhancer regions in mTEC II cells are significantly enriched ($P = 10^{-109}$; binomial test) for a binding-motif signature of the NF-κB family, correlating with specific expression of the *Nfkb2* gene in mTEC II cells (Fig. 2d–f and Extended Data Fig. 5). By contrast, mTEC-IV-specific enhancers were significantly enriched ($P = 10^{-805}$) for the POU class 2 transcription-factor motif, correlating with the specific expression of the *Pou2f3* gene in mTEC IV cells (Fig. 2d–f and Extended Data Fig. 5).

Fig. 2

Genetic and epigenetic characterization of TEC subsets.

a, Gating strategy for fluorescence-activated cell sorting to isolate mTEC subsets. **b**, qPCR analysis of mTEC subsets. Values represent fold change from the sample mean. $n = 3$ biologically independent animals. Data are mean (bars) and individual animals (dots). **c**, Normalized H3K4me2 profiles in 100-kb regions around differential mTEC

gene loci. Differential genes are specified by black arrows. Plots are representative of four animals from two independent experiments. **d**, Peak intensities of 2,302 differential peaks between the four mTEC subsets (clusters 1–7; *K*-means; $K = 7$) **e**, Transcription-factor motif analysis showing enrichment of motifs in accessible regions within H3K4me2-marked peaks; bars indicate motif abundance in H3K4me2 peaks clusters 1–7. False-discovery rate-corrected binomial test; $n = 137$ and 1,559 peaks. **f**, Projection of transcription factors onto the two-dimensional graph (Fig. 1).



AQ3

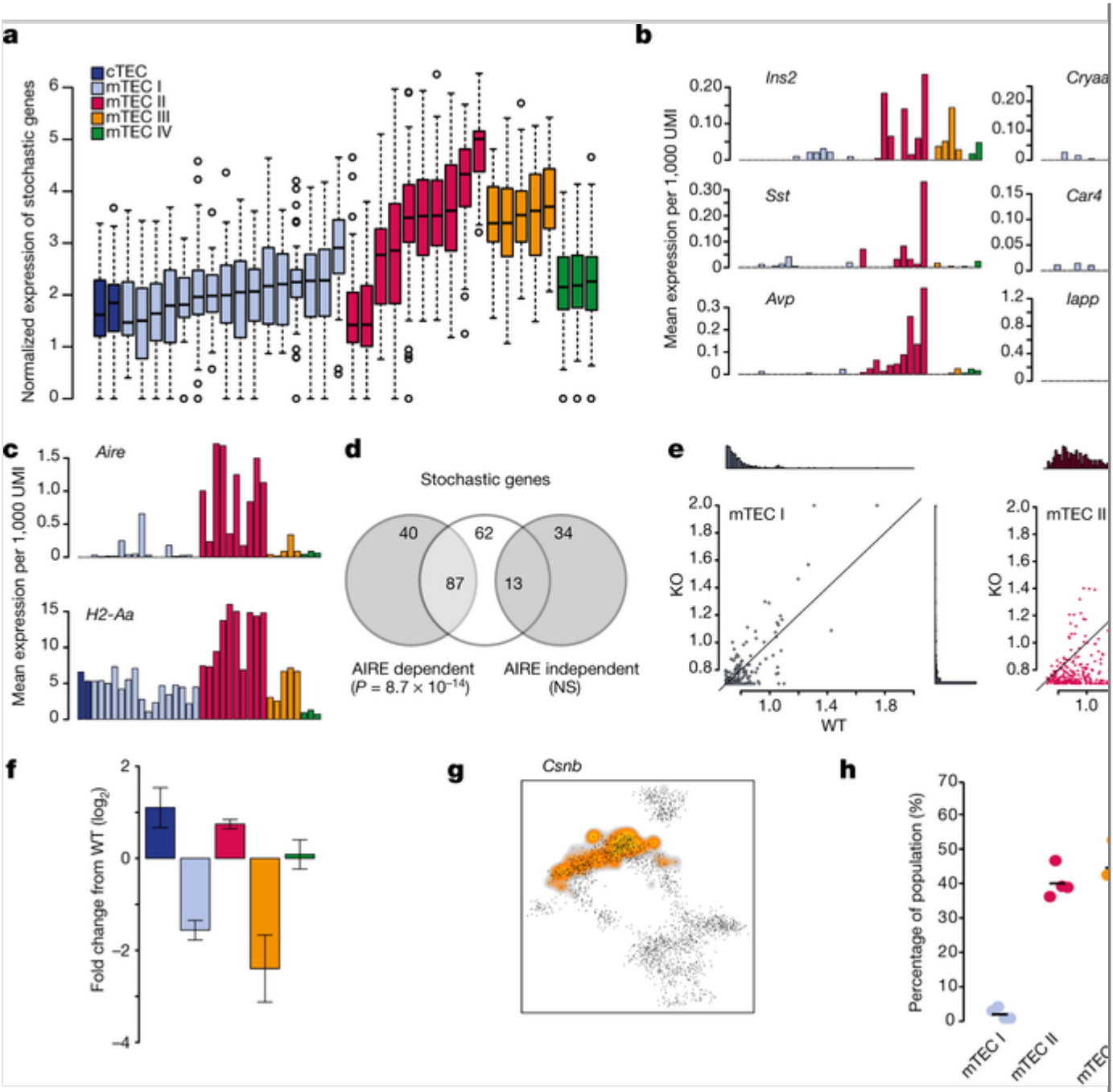
Given that one of the key functional roles of mTECs is to ectopically express and cross-present a plethora of tissue-restricted antigens (TRAs)[1, 2], we analysed the level of TRA gene expression within the TEC compartment. Because the expression of most TRAs is stochastic and AIRE-dependent[12, 13], we first defined a list of stochastically expressed genes in mTECs, based on their high expression variance and low correlation to other genes (Fig. 3, Extended Data Fig. 6, Supplementary Table 4 and Methods). As expected, mTEC II cells expressed the highest number of variable and uncorrelated genes (Fig. 3a, b and Extended Data Fig. 6). Notably,

mTEC III cells showed a high level of stochastic gene expression (Fig. 3a–c) in spite of low expression of AIRE. MARS-seq analysis of 1,332 $CD45^{-}EpCAM^{+}$ single cells from $Aire^{-/-}$ mice, validated that AIRE deficiency almost entirely eliminated the expression of the 'stochastic' genes within the mTEC II and III populations (87% and 67%, respectively; Fig. 3e and Extended Data Fig. 6). In addition, AIRE deficiency also resulted in a decrease in mTEC I and mTEC III populations (Fig. 3f). By contrast, mTEC II cells and cTECs, showed an increase in frequency, while the mTEC IV population was unaffected (Fig. 3f). In order to better understand the lineage relationships of the individual mTEC subsets we performed in vivo fate mapping using $Csnb^{cre}Rosa26^{tdTomato}$ reporter mice (Extended Data Fig. 6 and Methods). Whereas *Csnb* expression was restricted to the mTEC II and III subsets (Fig. 3g and Extended Data Fig. 6), the tdTomato reporter was expressed in mTEC II, III and IV, but was absent from mTEC I cells (Fig. 3h and Extended Data Fig. 6). This suggests that while mTEC IV cells may be developmentally derived from the $Csnb^{+}$ mTEC II and/or III populations, or from a common ancestor, the mTEC I population is not.

Fig. 3

Characterization of AIRE-dependent mTEC subsets.

a, Total expression distribution of stochastically expressed genes across the TEC metacells. In the box plots, bars indicate median, boxes are the first–third quantiles, whiskers, 5th–95th percentile and outlier are shown as circles. $n = 2{,}341$ single cells. **b**, **c**, Mean normalized expression of representative TRA genes (**b**), *Aire* and *H2-Aa* (**c**) across TEC metacells. **d**, Venn diagram depicting overlap between stochastically expressed genes and an established list of AIRE-dependent and -independent TRA genes. Hypergeometric test. $n = 388$ differentially expressed genes. **e**, Comparison of stochastic gene expression between *Aire* knockout (KO) and wild-type (WT) cells in TEC subsets. Marginal distributions are shown as histograms. Axes represent UMI count per 1,000 UMI, normalized to cell numbers. **f**, Bar plots showing $\log_2$ fold change between TEC subpopulation abundances in *Aire* knockout and wild-type mice. Error bars represent 95% confidence intervals. $n = 1{,}332$ knockout and 1,638 wild-type cells. **g**, Projection of *Csnb* onto the two-dimensional graph (Fig. 1). **h**, Percentage of tdTomato-expressing cells in mTEC I–IV. $n = 4$ biologically independent mice. Data are mean (line) and individual animals (dots).



Molecularly, the mTEC IV population is distinct from the other mTEC subsets, including the chromatin state, gene-expression profile and lack of stochastic gene expression (Figs. 1–3). On the basis of these data, we hypothesized that it may have a different functional role. This was further supported by specific expression of several genes that are associated with a rare epithelial lineage that is found in the gut, known as tuft (brush) cells[14, 15, 16] (Extended Data Fig. 7). In order to validate whether mTEC IV cells represent a putative tuft cell type, we compared their transcriptional profile to intestinal tuft cells and to the different TEC populations. Notably, mTEC IV cells were more similar to intestinal *Hpgds*-tdTomato^{+} tuft cells than to any of the TEC subpopulations (Fig. 4a). Specifically,

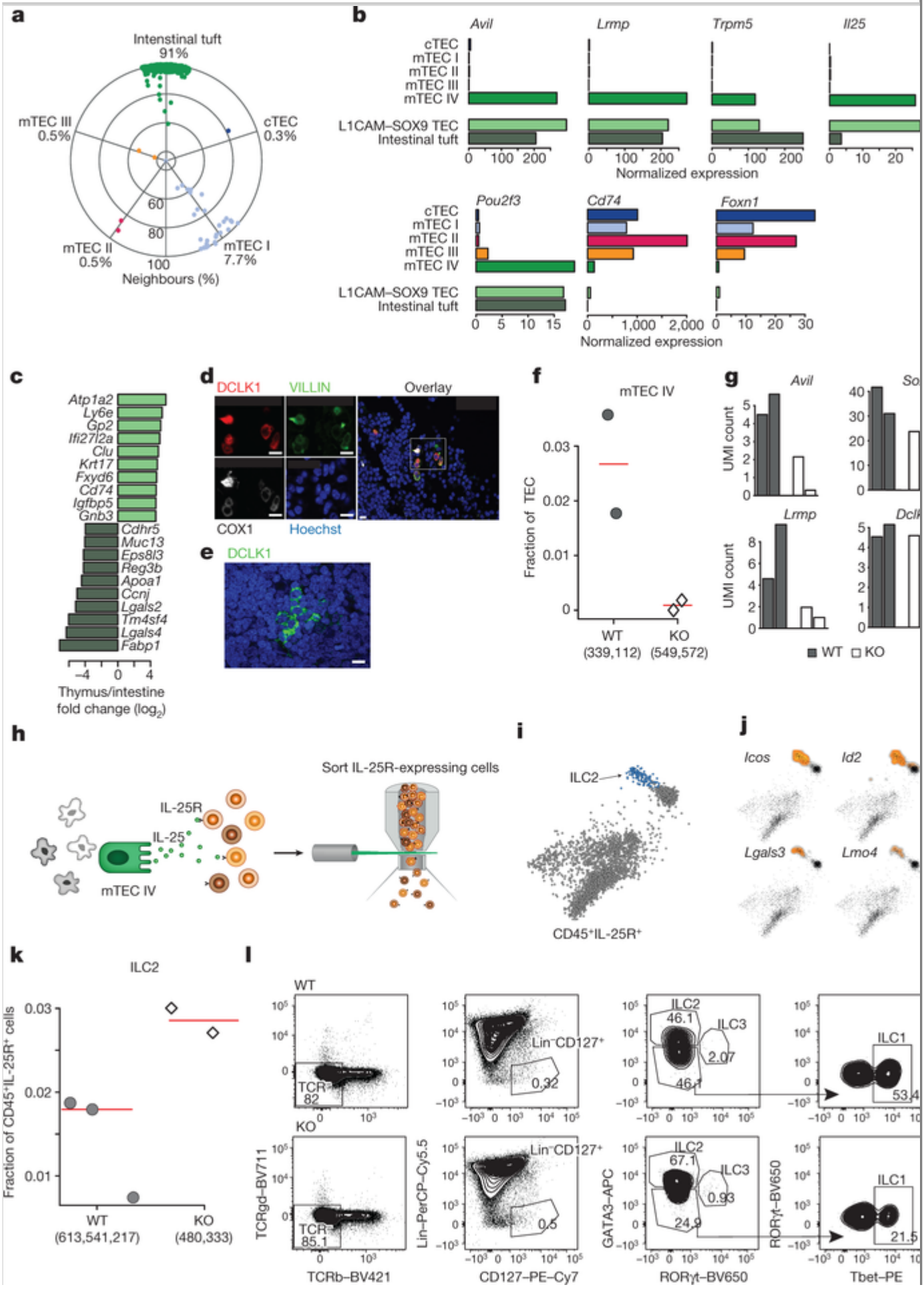
mTEC IV cells and intestinal tuft cells shared a large number of regulatory factors and tuft cell-specific genes including *Avil*, *Il25* and *Pou2f3*. However, they also showed differential gene expression, including for *Gp2* and *Gnb3*, which were only expressed in mTEC IV cells (Fig. 4b, c and Extended Data Fig. 7). Moreover, microscopy analysis of mTEC IV cells showed a typical tuft-like staining pattern in samples from both mice and humans (Fig. 4d, e). Finally, deficiency of a master regulator of intestinal tuft cells, *Pou2f3*, resulted in the complete loss of the mTEC IV population without affecting the development of any other TEC population (Fig. 4f, g and Extended Data Fig. 8), suggesting that they represent a bona fide tuft cell population.

Fig. 4

mTEC IV, a new TEC population with tuft-cell characteristics.

a, *K*-nearest neighbours quantification ($K = 50$) of thymic L1CAM$^+$*Sox9*-eGFP$^+$ cells (mTEC IV) mapped to both TEC subsets and intestinal *Hpgds*-tdTomato$^+$ tuft cells. Radial location signifies populations with the highest number of neighbours; *y* axis indicates percentage of all nearest neighbours. **b**, Normalized mean expression of differentially expressed genes across TEC subsets, L1CAM$^+$*Sox9*-eGFP$^+$ ($n = 376$) and intestinal *Hpgds*-tdTomato$^+$ ($n = 1{,}879$ single cells) cells. **c**, Differential gene expression between L1CAM$^+$*Sox9*-eGFP$^+$ and intestinal *Hpgds*$^-$tdTomato$^+$ cells ($\log_2$ fold change). **d**, **e**, Representative immunofluorescence imaging of tuft markers in thymic medulla sections from adult mice (**d**) and human tissue (**e**). Blue, DAPI. Scale bar = 10 μm. Images are representative of two independent experiments with similar results. **f**, Fraction of thymic mTEC IV cells from total TEC numbers in *Pou2f3* knockout and wild-type mice. Numbers in brackets indicate analysed cells in biological duplicates, indicated by individual circles and diamonds. Horizontal lines represent mean value. **g**, Mean expression of tuft markers in mTEC IV cells from wild-type and *Pou2f3* knockout mice. **h**, Experimental flow for exploring effect of mTEC IV cells on IL-25R$^+$ immune cells. **i**, Two-dimensional graphical representation of 3,500 CD45$^+$IL-25R$^+$ single cells separated into metacells. Blue dots label ILC2s. **j**, Projection of ILC2 markers onto the two-dimensional graph. **k**, Fraction of thymic ILC2 cluster from total CD45$^+$IL-25R$^+$ numbers in *Pou2f3* knockout and wild-type mice. Numbers in brackets indicate analysed cells in each replicate. Horizontal lines represent mean value and circles and diamonds indicate individual mice. **l**, ILC subtypes in *Pou2f3* knockout and wild-type thymi assessed by flow cytometry. Representative plots are shown ($n = 4$ biologically independent animals).

AQ5


a
Intenstinal tuft
91%
mTEC III
0.5%
cTEC
0.3%
mTEC II
0.5%
mTEC I
7.7%
60
80
100
Neighbours (%)
b
Avil
Lrmp
Trpm5
Il25
cTEC
mTEC I
mTEC II
mTEC III
mTEC IV
L1CAM–SOX9 TEC
Intestinal tuft
Normalized expression
Pou2f3
Cd74
Foxn1
c
Atp1a2
Ly6e
Gp2
Ifi27l2a
Clu
Krt17
Fxyd6
Cd74
Igfbp5
Gnb3
Cdhr5
Muc13
Eps8l3
Reg3b
Apoa1
Ccnj
Lgals2
Tm4sf4
Lgals4
Fabp1
Thymus/intestine
fold change (log2)
d
DCLK1
VILLIN
Overlay
COX1
Hoechst
e
DCLK1
f
mTEC IV
Fraction of TEC
WT
(339,112)
KO
(549,572)
g
Avil
Lrmp
UMI count
WT
KO
h
Sort IL-25R-expressing cells
IL-25R
IL-25
mTEC IV
i
ILC2
CD45+IL-25R+
j
Icos
Id2
Lgals3
Lmo4
k
ILC2
Fraction of CD45+IL-25R+ cells
WT
(613,541,217)
KO
(480,333)
l
WT
KO
TCR
82
85.1
Lin−CD127+
0.32
0.5
ILC2
46.1
ILC3
2.07
ILC1
53.4
67.1
0.93
24.9
21.5
TCRgd–BV711
TCRb–BV421
Lin–PerCP–Cy5.5
CD127–PE–Cy7
GATA3–APC
RORγt–BV650
Tbet–PE

AQ4

Because the mTEC IV population is characterized by high and exclusive expression of IL-25 in the stroma of the thymus (Fig. 4b), we next studied their potential impact on thymic cells expressing the IL-25 receptor (*IL17rb*) (Fig. 4h). To this end, we first characterized 3,500 thymic $CD45^{+}IL\text{-}25R^{+}$ cells using MARS-seq and found five different subpopulations with distinct transcriptional states, including several $CD3^{+}$ thymocyte subsets, as well as thymus-resident type-2 innate lymphoid cells (ILC2s) (Fig. 4h–j and Extended Data Fig. 9). Notably, the loss of mTEC IV cells in *Pou2f3*$^{-/-}$ mice was accompanied by a significant increase in the thymic ILC2 population (Fig. 4k), whereas it had no significant impact on other thymic $CD45^{+}IL\text{-}25R^{+}$ cells or the main T cell subsets (Extended Data Fig. 9). The increase in the ILC2 ($Lin^{-}TCR^{-}CD127^{high}Tbet^{-}ROR\gamma t^{-}GATA3^{+}$) subset in *Pou2f3*$^{-/-}$ versus wild-type mice was further confirmed by conventional flow cytometry (Fig. 4l and Extended Data Fig. 9).

In summary, our study provides a comprehensive atlas of the stromal populations in the thymus of adult mice and during differentiation. We uncover unexpected complexity and diversity in this compartment, including an mTEC population with tuft-cell characteristics. This study also highlights important qualitative differences between the embryonic and postnatal thymus, which have, thus far, remained uncharacterized. As such, it clarifies many past controversies and confusions about the molecular, developmental and functional characteristics of the previously characterized TEC subsets and TRA expression, enabling the field to progress forward on a stable and common molecular blueprint. Nevertheless, several important open questions remain to be addressed in the future, including other functions of the mTEC IV population in the thymus and the lineage relationship and origins of the individual TEC subsets. Specifically, whether the mTEC II, III and IV populations are derived from a common ancestor cell, or whether mTEC IV cells are derived from the mTEC II and/or III populations.

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Mice

All mice were maintained under specific pathogen-free conditions at the Weizmann Institute's animal facility and were handled in accordance with the guidelines of the Institutional Animal Care and Use Committee (25720316-2). Wild-type C57BL/6 (B6) mice were purchased from Harlan Laboratories. *Aire* knockout C57BL/6 mice were purchased from Jackson Laboratories. $Sox9^{eGFP}$ mice were generated as described previously[17]. *Pou2f3* knockout and $Hpgds^{tdTomato}$ mice were maintained under pathogen-free conditions at the IGMM and IGF animal facilities (RAM). $Csnb^{cre}$ mice were generated using bacterial artificial chromosome (BAC) recombination. In brief, homology arms corresponding to a sequence downstream to the ATG of the *Csnb* gene were cloned into a Cre-recombinase-coding plasmid (pConst-Cre; provided by G. Schutz), followed by homologous recombination of the *cre* coding sequence into the BAC containing the *Csnb* gene (RP23-218H23, CHORI). Flp recombination was then used to remove the antibiotic resistance cassette. The final BAC was injected into fertilized BALB/c oocytes by the Weizmann transgenic core facility. Heterozygous mutants were backcrossed to C57BL/6 mice. $Csnb^{cre}Rosa26^{tdTomato}$ reporter mice were generated by breeding heterozygous $Csnb^{cre}$ mice with $Rosa26^{tdTomato}$ mice (Jackson Laboratories).

Human samples

Human thymus samples were obtained during the course of corrective cardiac surgery at Schneider Children's Medical Center of Israel, following ethical approval (0781-16-RMC). An informed consent was signed by parents of the patients before obtaining the thymus.

Ethical compliance

All animals were housed according to guidelines at the Weizmann Institute of Science and the IGMM and IGF animal facilities (RAM). All experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC), application number 25720316-2.

All human experiments were done according to guidelines at the Weizmann Institute of Science and approved by Helsinki ethics committee, approval number 0781-16-RMC.

Isolation of mouse thymic stromal cells

Thymi from 4–8-week-old mice (unless otherwise stated, for embryonic or neonatal thymi) were placed into cold 1× PBS supplemented with 2% fetal bovine serum (FBS, Invitrogen). Thymi were chopped into small pieces and disintegrated by enzymatic digestion for 30–40 min in a 37 °C water bath, using 0.3 mg ml^{-1} collagenase D (Roche, 1088858), 1 mg ml^{-1} dispase II (Roche, 04942078001) and 10 ng ml^{-1} DNase I (Sigma-Aldrich, DN25) in RPMI supplemented with 2% FBS. Cells were then filtered through a 50-μm mesh filter and washed with 5–10 ml MACS buffer (1× PBS with 5 mM EDTA and 2% FBS), followed by centrifugation at 230*g* for 4 min. Percoll gradient density centrifugation was performed in order to enrich the stromal compartment. In brief, cells were resuspended in 2 ml of 1.115 g ml^{-1} isotonic Percoll (Sigma-Aldrich, P1644) and placed at the bottom of a tube. Subsequently, 1 ml of isotonic 1.065 g ml^{-1} Percoll and then 1 ml of 1× PBS were layered on top. The Percoll gradient was centrifuged at 2,700 r.p.m., 4 °C, with no deceleration for 30 min. The thymic stroma accumulated between the top and middle layers, and was collected and washed with MACS buffer and centrifuged at 230*g* for 4 min. Embryonic thymi were not subjected to Percoll separation, but were treated with red blood cell lysis buffer (15 mM ammonium chloride, 1 mM potassium bicarbonate, 10 μM EDTA in double distilled water, pH 7.3).

Isolation of thymic haematopoietic cells

Thymi from 6–12-week-old mice were surgically removed and placed in PBS on ice. Thymi were trimmed of fat and connective tissues, and thymocytes were extracted into a single-cell suspension by pressing the thymic lobes against a 70-μm cell strainer. Cells were washed in MACS buffer (1× PBS with 2% FBS and 5 mM EDTA pH 8.0).

Isolation of intestinal epithelial cells

Mouse small intestines were isolated, flushed with PBS and incised along their length. The tissue was incubated in 30 mM EDTA (Sigma-Aldrich) in HBSS pH 7.4 (Life Technologies) on ice, and transferred to DMEM (Life Technologies) supplemented with 10% FBS (Sigma-Aldrich). Vigorous shaking yielded the epithelial fraction that was then incubated with 100 μl of dispase (BD Biosciences) in 10 ml of HBSS, supplemented with 100 μl of DNase I at 2,000 Kunitz (Sigma). The single-cell preparation was obtained by filtration through a 30-μm mesh and used for further staining.

Antibodies and reagents for flow cytometry

APC–Cy7–EpCAM (118218), PE–Ly51 (108308), pacific blue–I-A/I-E (107620), PerCP–Cy5.5–CD45 (103132), FITC–CD34 (343603), APC–CD31 (102409), PE–Cy7–CD45 (103114), FITC–ITGB4 (123605), PE–ITGB4 (123610) and PE–IL-17RB (IL-25R,146305) were purchased from Biolegend; APC–L1CAM (FAB5674R) was purchased from Novus; APC–GATA3 (560078), brilliant violet 650–RORγt (564722), biotinylated CD3 (553060), biotinylated CD4 (553728), biotinylated CD8 (553029), biotinylated B220 (553086), biotinylated Ter119 (553672), biotinylated CD11c (553800), biotinylated Gr1 (553125), FITC–CD45 (553080), FITC–CD8 (553031), APC–Ter119 (557909), PE–Cy7–CD45 (552848), brilliant violet 711–CD4 (563726), APC–CD8 (553035), horizon V500–CD44 (560780), brilliant violet 711–Tcrgd (563994) and horizon V450–Tcrb (560706) were purchased from BD; and PerCP–eFluor710–Ly6d (46-5974-80), PE–Cy7–CD127 (25-1273-82), PE–Tbet (12-5825-82), eFluor780–CD4 (47-0042-82) and eFluor780–CD25 (47-025182) were purchased from eBioscience. The following materials were also used for FACS staining: 7-AAD Viability Staining Solution (Biolegend, 420404); Biotinylated Ulex Europaeus Agglutinin I (UEA I) (Vector laboratories, B-1065), PE–Cy7–streptavidin (405206, Biolegend), PerCP–Cy5.5–streptavidin (554064, BD), SYTOX (S34857, Invitrogen), FcR blocker (BE0307, Bio X Cell).

Flow cytometry and sorting

Cells were stained in MACS buffer (1× PBS with 2% FBS and 5 mM EDTA pH 8.0) with specific antibodies for 30 min at 4 °C. For haematopoietic cells, to avoid non-specific binding of antibodies to Fcγ receptors, anti-mouse CD16/CD32 monoclonal antibody (Bio X cell, BE0307 or Biolegend, 10130) was added to the antibody mix. Following staining, cells were washed and resuspended in MACS buffer. Secondary staining with streptavidin was performed in a similar manner. After the wash, cells intended for intracellular staining were fixed and permeabilized using the eBioscience fixation/permeabilization kit according to kit instructions, and stained with antibodies of intracellular markers for 2 h at 4 °C. Cells were sorted on a BD FACSAria Special Order Research Product (SORP) or BD FACSAria Fusion/II or III, or analysed on a BD LSRFortessa or BD FACSCantoII. Spectral overlap between fluorescent dyes was compensated using single-stained controls. Pre-gating was first done for live cells (in non-fixed samples) based on a 7-AAD, DAPI or SYTOX stain, followed by single-cell gating according to the FSC-A versus FCS-W plot. Data analysis was performed using FlowJo software (Tree Star Inc.).

mTEC subsets (Fig. 2a) were sorted from $CD45^{-}EpCAM^{+}$ cells using the following sub-sorting gates: mTEC I: MHC-$II^{low}ITGB4^{+}L1CAM^{-}$; mTEC II: MHC-$II^{+}Ly6d^{-}$; mTEC III: $ITGB4^{-}L1CAM^{-}Ly6d^{+}$; mTEC IV: MHC-$II^{low}L1CAM^{+}$.

Single-cell index sorting

Isolated cells were single-cell sorted into 384-well cell capture plates containing 2 µl of lysis solution and barcoded poly(T) reverse-transcription primers for single-cell RNA-seq. Barcoded single-cell capture plates were prepared with a Bravo automated liquid handling platform (Agilent), as described previously[4]. To record marker levels of each single cell, the FACS Diva 7 ‘index sorting’ function was activated during single-cell sorting.

Library preparation for single-cell RNA sequencing

Single-cell libraries were prepared as previously described[4]. In brief, 384-well plates, which contained lysis buffer and barcoded reverse-transcription poly-T primers, were immediately spun down and placed on dry ice. For barcoding and reverse transcription, a reverse-transcription reaction mix was added and the plates were placed in a PCR machine set to the appropriate program. All barcoded samples were pooled, followed by addition of exonuclease to remove excess RT–PCR primers. A purification step using SPRI beads that bind cDNA and RNA was performed after this step, as well as after each of the following steps. The pooled single-stranded cDNA was converted to a double-stranded DNA using a designated kit, in order to perform in vitro transcription of RNA molecules. The template DNA was then removed using DNase, and the generated RNA was fragmented and ligated to barcoded Illumina adapters. Reverse transcription of this ligation product was done using primers specific for the Illumina adapters, and libraries of the resulting cDNA were generated and enriched by 12–15 PCR cycles.

Low-level processing and filtering

All RNA-seq libraries (pooled at equimolar concentration) were sequenced using the Illumina NextSeq 500 at a median sequencing depth of 16,289 reads per single cell. Sequences were mapped to mouse genome (mm9), demultiplexed and filtered as previously described[4, 18], extracting a set of unique molecular identifiers (UMI) that define distinct transcripts in single cells for further processing. We estimated the level of spurious UMIs in the data using statistics on empty MARS-seq wells (median noise 2.8%; Extended Data Fig. 2). Mapping of reads was done using HISAT (version 0.1.6)[19]; reads with multiple mapping positions were

excluded. Reads were associated with genes if they were mapped to an exon, using the UCSC genome browser as reference. Exons of different genes that shared genomic position on the same strand were considered a single gene with a concatenated gene symbol. Cells with less than 500 UMIs were discarded from the analysis. After filtering, cells contained a median of 1,711 unique molecules per cell.

Data processing and clustering

The MetaCell pipeline was used to derive informative genes and compute cell-to-cell similarity, to compute KNN graph covers and derive distribution of RNA in cohesive groups of cells (or metacells), and to derive strongly separated clusters using bootstrap analysis and computation of graph covers on resampled data. The MetaCell package is described in detail in Supplementary Note 1. Default parameters were used unless otherwise stated.

Clustering for Figs. 1, 2 was done on a combined set of cells from two sources: (1) 1,972 $CD45^-$ thymic cells and (2) 1,542 $CD45^-EpCAM^+$ thymic cells (Extended Data Fig. 2). Clustering resulted in 49 clusters. Clusters with increased expression of *Hbb-b1*, *Trbc2* or *C1qb*, which are markers for red blood cells, T cells or macrophages, respectively (mean expression >10 times the median across clusters), were marked as contaminants and discarded from further analysis (Extended Data Fig. 3). We performed hierarchical clustering over the clusters structure and divided clusters into epithelial, endothelial and fibroblast groups by cutree.

Mapping cells to an existing cluster model

Given an existing reference single-cell dataset and cluster model, and a new set of single-cell profiles, we extracted for each new cell the K ($K = 10$) reference cells with top Pearson correlation on transformed marker gene UMIs as described above. The distribution of cluster memberships over these K-neighbours was used to define the new cell reference cluster (by majority voting), and was applied for visualizing new cells by weighted average of the x and y coordinates of the clusters.

Clustering of development TEC and comparison to the existing model

Clustering of TEC during embryonic and postnatal development (Fig. 3) was done on a combined dataset of 1,343 E14.5, 895 E18.5 and 836 6 days postnatal epithelial cells. Two-dimensional projection of the resulting clustering was produced by MetaCell. However, in order to maintain the structure of the two-

dimensional projection in Fig. 2, while enabling the discovery of new transcriptional states, we computed the KNN structure of the combined developmental and mature TEC. Developmental clusters with more than 20% cells for which >20% of their neighbours are within the mature dataset were associated with the mature projection, and their two-dimensional coordinates were determined by their mature neighbours. All other developmental clusters were assigned their regular coordinates (Fig. 3a and Extended Data Fig. 6).

For further refinement of the early TEC population (as in Fig. 3g), all developmental metacells with mean expression >0.25 times the median across clusters were included.

Gene modules and cell cycle analysis

Identification of ribosomal, cell cycle or other broadly expressed gene modules was done by clustering genes in a downsampled UMI matrix (500 molecules per cell). We filtered genes with total molecule (UMI) count lower than 5 and variance-to-mean ratio lower than 1.2. Hierarchical clustering using Ward's method was performed for detecting 30 fine-grained clusters. After removing clusters with mean Pearson intracorrelation lower than 0.025, 26 gene modules were retained. Manual annotation of the gene clusters was performed. This resulted in the identification of 26 modules with 8–136 genes, among which the cell cycle module contained 73 genes. Ribosomal modules (86 genes) were excluded from clustering analysis. Expression of cell cycle genes is a good indicator of proliferation[20]. To determine proliferation status of genes, we examined the pooled normalized expression of the cell cycle module across genes. This measurement showed a bimodal distribution, correlating with the total UMI count of cells (Extended Data Fig. 4e).

GO enrichment

Gene enrichment analysis was done using metascape software with mouse whole-genome backup.

Analysis of stochastically expressed genes

In order to define stochastically expressed genes, genes with less than 50 total UMIs were discarded (list A). Of this list, genes for which the corrected variance was greater than 1 (by a linear fit of the variance to the mean expression) were defined as variably expressed (list B). Gene-to-gene Pearson correlations were then computed on the UMI values of genes from list B (normalized to cell size). In order to discard tightly correlated gene expression programs, only genes for which the

third highest correlation with other genes was less than 0.25 were considered as stochastically expressed (list C; Supplementary Table 4).

An established list of TRA genes was taken from a previous publication[21]. AIRE dependency of these TRA was determined by a twofold reduction in *Aire* knockout mice measured in the previous study[21]. Both AIRE-dependent and -independent lists were intersected with list B for further analysis.

Immunofluorescence staining of frozen thymic sections

Thymi from 4–8-week-old female mice were embedded in OCT compound (Tissue-Tek, Sakura) and frozen on dry ice. Cryostat sections (6 µm) were fixed with ice-cold acetone for 10 min and incubated with primary antibody (anti-AIRE AF488-conjugated, 04-150; Millipore) diluted in 1% BSA in PBS for 60 min at room temperature. Sections were washed three times with PBS and incubated with DAPI staining for 10 min at room temperature, followed by three washes with PBS.

Immunoflourescence staining of PFA-fixed frozen thymic sections

Thymi from 4–8-week-old wild-type, $Sox9^{eGFP}$, $Csnb^{cre+}Rosa26^{tdTomato}$ and $Csnb^{cre-}Rosa26^{tdTomato}$ mice were isolated, cleaned and fixed for 3 h with ice-cold 3.7% formaldehyde, followed by overnight incubation with cryoprotection solution (3.7% formaldehyde, 30% sucrose in 1× PBS). Thymi were embedded in OCT and frozen on dry ice. Cryostat sections (6–7 µm) were permeabilized and blocked with blocking buffer (TBS, pH 7.4, 5% goat serum or BSA, and 0.1% Triton X-100) for 30 min at room temperature. Sections were then incubated with primary antibodies diluted in blocking buffer overnight at 4 °C. Primary antibodies used were as follows: anti-PSMB11 (pd021, MBL); anti-PIGR (AF2800, R&D Systems); anti-DCLK1 (ab37994, Abcam); APC-conjugated anti-ITGA6 (313615; Biolegend). Sections were washed twice with TBST (1× TBS supplemented with 0.1% Tween-20). Sections stained with unconjugated antibodies were incubated with a secondary antibody (goat anti-rabbit AF555, Jackson Laboratories) for 60 min at room temperature. DAPI staining was performed for 10 min at room temperature.

Immunoflourescence staining of paraffin-embedded thymic sections

Thymi from 4–8-week-old mice or 8-day-old male humans were fixed in 4% PFA for 48 h, followed by embedding in paraffin. Subsequently, 5-µm-thick sections were dewaxed in xylene and rehydrated in graded alcohol baths. Antigen retrieval

was performed by boiling slides for 20 min in 10 mM sodium citrate buffer, pH 6.0. Nonspecific binding sites were blocked in blocking buffer (TBS pH 7.4, 5% goat serum and 0.1% Triton X-100) for 30 min at room temperature. Sections were incubated with primary antibodies diluted in blocking buffer overnight at 4 °C. Primary antibodies used were as follows: anti-villin (MAB1671, Millipore); anti-COX1 (sc-1754, Santa Cruz Biotechnology); anti-DCLK1 (ab37994, Abcam). Slides were washed two times with TBST before incubation with fluorescent secondary antibodies conjugated to AF488, AF555, Cy3 or Cy5 (Jackson ImmunoResearch Laboratories) and Hoechst (Sigma-Aldrich) or DAPI in TBS–0.1% Triton X-100.

Single-molecule RNA fluorescence in situ hybridization (FISH)

Single-molecule FISH probe libraries consisting of 48 probes with a length of 20 bp were designed as previously described[22], constructed and provided by Agentek. The following probe libraries were used: *Avil*, *Sbsn* coupled to Cy5, *Epcam* and *Pigr* coupled to AF594. RNA FISH was performed as previously described. In brief, thymi from 4–8-week-old female mice were isolated, cleaned and fixed for 3 h with ice-cold 3.7% formaldehyde followed by overnight incubation with cryoprotection solution (3.7% formaldehyde, 30% sucrose in PBS). Thymi were embedded in OCT and frozen on dry ice. Cryostat sections (5–8 μm) were air-dried and fixed again with 3.7% formaldehyde for 5 min following by 2 h incubation with 70% ethanol at 4 °C. Sections were rehydrated with 2× SSC, and treated with proteinase K in 2× SSC for 10 min at room temperature. Hybridization was performed overnight by 20% formamide in 2× SSC with 0.1 ng μl^{-1} of the desired probes at 30 °C. DAPI (to stain the nuclei) was added during the washes, and sections were incubated with fresh GLOX buffer (10 mM Tris pH 8.0 and 0.4% glucose in 2× SSC).

Imaging and image analysis

Imaging was performed on Ultima Multiphoton Microscope, Nikon Eclipse TI-S fluorescence microscope, or Nikon-Ti-E inverted fluorescence microscope with 60× and 100× oil-immersion objectives and a Photometrics Pixis 1024 CCD camera, using MetaMorph software (Molecular Devices) or NIS element software (Nikon). Image analysis was performed with ImageJ software.

Indexing first chromatin immunoprecipitation followed by deep sequencing (iChIP–seq)

iChIP–seq was prepared as previously described[10]. In brief, cells were cross-linked for 8 min in 1% formaldehyde and quenched for 5 min in 0.125 M glycine before sorting. Cells were sorted using the described sorting strategy and frozen. Cell pellets were lysed in 0.5% SDS and sheared with the NGS Bioruptor Sonicator (Diagenode). Sheared chromatin was immobilized on 15 µl Dynabeads Protein G (Invitrogen) with 1.3 µg of anti-H3 antibody (Abcam). Magnetized chromatin was then washed with 10 mM Tris-HCl supplemented with 1× PI. Chromatin was end-repaired, dA-tailed and ligated with sequencing adapters containing Illumina P5 and P7 sequences. Indexed chromatin was pooled and incubated with 2.5 µg H3K4me2 antibody (ab32356, Abcam) at 4 °C for 3 h and for an additional hour with protein G magnetic beads (Invitrogen). Magnetized chromatin was washed and reverse cross-linked. DNA was subsequently purified with 1.65× SPRI and amplified by PCR with 0.5 µM of forward and reverse primers containing Illuminia P5-rd1 and P7-rd2 sequences. Library concentration was measured with a Qubit fluorometer and mean molecule size was determined by TapeStation (Agilent). DNA libraries were sequenced on an Illumina NextSeq 500 with an average of over 10 million aligned reads per replicate.

ATAC-seq

To profile open chromatin, we used an assay of transposase-accessible chromatin following sequencing (ATAC-seq) as previously published[23], with modifications as previously described[10]. In brief, cell populations were sorted in 400 µl of MACS buffer (1× PBS, 0.5% BSA, 2 mM EDTA) and pelleted by centrifugation for 15 min at 500*g* and 4 °C with low acceleration and brake settings. Cell pellets were washed once with 1× PBS and cells were pelleted by centrifugation using the previous settings. Cell pellets were resuspended in 25 µl of lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Igepal CA-630) and nuclei were pelleted by centrifugation for 30 min at 500*g*, 4 °C with low acceleration and brake settings. Supernatant was discarded and nuclei were resuspended in 25 µl reaction buffer containing 2 µl of Tn5 transposase and 12.5 µl of TD buffer (Nextera Sample preparation kit from Illumina). The reaction was incubated at 37 °C for 1 h. Then, 5 µl of clean-up buffer (900 mM NaCl, 300 mM EDTA), 2 µl of 5% SDS and 2 µl of Proteinase K (NEB) were added and incubated for 30 min at 4 °C. Tagmentated DNA was isolated using 2× SPRI beads clean-up. For library amplification, two sequential nine-cycle PCR runs were performed in order to enrich small tagmentated DNA fragments. We used 2 µl of indexing primers included in the Nextera Index kit and KAPA HiFi HotStart ready mix. After the first PCR, the libraries were selected for small fragments (less than 600 bp) using

SPRI clean-up. Then a second PCR was performed with the same conditions in order to obtain the final library.

Processing of iChIP–seq, ATAC-seq and chromatin peak calling

Reads were aligned to the mouse reference genome (mm9, NCBI v.37) using Bowtie aligner version 1.0.0[24] with best match parameters (bowtie -m 1–sam–best–strata -v 2). To identify regions of enrichment (peaks) from ChIP–seq reads of H3K4me2, we used the HOMER package makeTagDirectory followed by the findPeaks command with the histone parameter[25] and IDR filtering for reproducible peaks across replicates[26]. Union peaks file were generated by combining and merging overlapping peaks in all samples.

Chromatin analysis

For clustering of differential peaks, we first averaged peak sizes across replicates. We defined differential peaks as peaks for which the maximum value is more than fourfold higher than their minimum value. We then normalized peaks intensities and performed *K* means clustering ($K = 7$). For motif finding, we independently called peaks in ATAC-seq, as above, and identified the maximum peak that overlapped each H3K4me2 region. The overlapping sequences were input for HOMER package motif finder algorithm findMotifGenome[25].

Gene tracks and visualization

All gene tracks were visualized as bigWig files of the combined replicates normalized to 10,000,000 reads, using Integrative Genomics Viewer (http://www.broadinstitute.org/igv).

Real-time PCR analysis

Cells were sorted into 40–50 µl of lysis/binding buffer (Life Technologies). mRNA was captured with 15 µl of Dynabeads oligo(dT) (Life Technologies), washed and eluted at 85 °C with 10 µl of 10 mM Tris-Cl (pH 7.5). The purified RNA was used for cDNA synthesis using the High-Capacity cDNA Reverse-Transcription kit (Applied Biosystems) and polyT primers. The subsequent qPCR analysis was performed using the Fast SYBR Green Master Mix (Life technologies). Differential expression was calculated according to the $\Delta\Delta C_t$ method. Specific qPCR primers: *Actb*: GGAGGGGGTTGAGGTGTT, TGTGCACTTTTATTGGTCTCAAG; *Sox4*: GCTGGGCTTTCTCCTCCT, AGGCTGGCCTGCTACTCC; *Aire*: TGGGCTGATTAGGACCAAGA, ACAAAGATCAGGGCCATCTG; *Avil*:

GCATCAGGACCCACATCTGC, ATGCTGTGGCACATGGTAGAC; *Sbsn*: CACCATGCCCTAAACTGATGC, ACAAAGCTCAAAGCAGCCCTC; *cre*: AGGAGAATGTGGATGCTGGGG, CAATTTCGGCAATGCGCAGC; *Csnb*: AAACTTCAGAAGGTGAATCTCATGG, GCTGGATGTTTTGTGGGACG.

Lists of ribosomal and cell cycle modules

Ribosomal genes. 2810422J05Rik, AC151602.1, AL663027.1, Cfl1, Eef1b2, Ftl1, Gm10059, Gm10076, Gm10443, Gm11361, Gm11808, Gm11942, Gm12630, Gm13408, Gm14456, Gm15427, Gm15459, Gm15710, Gm3788, Gm4149, Gm5244, Gm5559, Gm8730, Gm8759, Gnb2l1, Hspa8, Mif, Rpl10, Rpl13, Rpl13a, Rpl14, Rpl18a, Rpl22, Rpl26, Rpl28-ps3, Rpl29, Rpl32, Rpl34, Rpl37, Rpl37a, Rpl38, Rpl38-ps2, Rpl4, Rpl7a-ps12, Rpl8, Rpl9-ps6, Rplp0, Rplp1, Rplp2, Rps10, Rps10-ps1, Rps14, Rps15, Rps18, Rps19, Rps2, Rps20, Rps21, Rps26, Rps28, Rps3, Rps3a, Rps5, Rps8, Rps9, Rpsa, Rpsa-ps10, Snord35a, Tmsb10, Tpt1, B2m, Eef1a1, Gas5, Gm13456, Gm15500, Gm16247, Rpl23, Rpl35a, Rpl7, Rps11, Rps24, Rps25, S100a11, Tmsb4x, Ubl5, mmu-mir-703.

Cell cycle genes. 1700003E16Rik, 2810417H13Rik, 2900006K08Rik, 4833427G06Rik, 4930473A06Rik, 5133401N09Rik, 6820408C15Rik, Arhgap11a, Arl6ip1, Birc5, Ccdc108, Ccdc113, Ccdc151, Ccdc17, Ccdc19, Ccdc30, Ccdc39, Ccdc40, Ccdc67, Ccno, Cdc20, Cdca8, Cenpf, Ckap2, Ckap2l, Cks1b, Dcdc2a, Dek, Dnahc9, Dnajb13, E030019B06Rik, Foxm1, Gm11423, Gm9938, H2afx, Hist1h1a, Hist1h1b, Hist1h1d, Hist1h1e, Hist1h2ac, Hist1h2ae, Hist1h2ao, Hist1h2ap, Hist1h3c, Hist1h3e, Hist1h4d, Hmgb2, Hnrnpa2b1, Ift46, Kif15, Kif24, Lrrc23, Lrrc46, Mki67, Phospho2, Pih1d2, Plk1, Rsph1, Rsph9, Smc4, Supt16h, Tekt1, Tekt4, Tmem107, Top2a, Ttll6, Tuba1b, Tubb2c, Tubb5, Ube2c, Wdr52, Wdr65, Zmynd10.

Reporting summary

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

Code availability

No custom software was used to collect data. ChIP–seq and ATAC-seq data analyses, including motif finding, were done with the HOMER package[25]. Single-cell data were analysed with the MetaCell package, which is available upon request.

Data availability

RNA, ChIP–seq and ATAC-seq data reported in this paper were deposited with Gene Expression Omnibus under accession numbers: GSE103967, GSE103968, GSE103969 and GSE103970.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0346-1.

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0346-1.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0346-1.

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

Acknowledgements

I.A. is supported by the Chan Zuckerberg Initiative, the HHMI International Scholar award, the European Research Council Consolidator Grant (ERC-COG)724471-HemTree2.0, the Israel Science Foundation (703/15), the Ernest and Bonnie Beutler Research Program of Excellence in Genomic Medicine, the Helen and Martin Kimmel award for innovative investigation, a Minerva Stiftung research grant, the Israeli Ministry of Science, Technology, and Space, the David and Fela Shapell Family Foundation, the NeuroMac DFG/Transregional Collaborative Research Center Grant, and the Abramson Family Center for Young Scientists. I.A. is the incumbent of the Alan and Laraine Fischer Career Development Chair. J.A. is supported by the ERC-2016-CoG-724821, Israel Science Foundation (1796/16 and 722/14), the Sy Syms Foundation; US–Israel Binational Foundation, Maurice and Vivienne Wohl Charitable Foundation; Goodman Family Charitable Lead Annuity Trust; Ruth and Samuel David Gameroff Family Foundation. J.A. is an incumbent of the Dr. Celia Zwillenberg-Fridman and Dr. Lutz Fridman Career Development Chair; A.G. is a recipient of the Clore fellowship; P.J. and N.T. are supported by ANR-17-CE15-TUFTEFF and the Labex EpiGenMed; M.P. is supported by the Labex EpiGenMed and the FRM; F.G. and V.Z. are supported by CNRS; and N.T.

and P.J. are supported by Inserm. We thank M. Boyer-Clavel and S. Gailhac of Montpellier Rio Imaging for cell sorting and the RAM animal facility of the IGMM.

Author contributions

C.B., S.N., A.G., N.K., J.A. and I.A. designed the project, planned the experiments and wrote the manuscript. C.B., S.N. and N.K. performed experiments, A.G. and E.D. analysed the data, M.P., F.G., A.M., N.T., P.J. and V.Z. performed intestinal tuft and *Pou2f3* knockout mouse experiments. A.C. generated the *Csnb*-reporter mouse. B.T. and S.I. contributed to the single-molecule RNA fluorescence in situ hybridization experiment. O.G. provided human samples.

Competing interests The authors declare no competing interests.

Extended data figures and tables

Extended Data Fig. 1

Single-cell data quality controls.

a, Summary of all single cells analysed in this study, divided into experimental procedures. ‘nbatches’ indicates number of technical replicates; ‘ncells’ indicates number of cells after filtering (see Methods). **b**–**e**, Colour-coded tracks summarizing the number of Illumina reads per cell (**b**), transcripts (UMI) detected in each cell (**c**), fraction of analysed cells from each amplification batch (**d**) and estimation of technical noise for each amplification batch (**e**). Cells are coloured by experimental procedure. Technical noise is assessed by genomic UMIs in empty wells as previously described[4] (see Methods).

Extended Data Fig. 2

Thymic stroma sorting and clustering.

a, Flow cytometry schematic of thymic cells showing isolation of stroma cells, as well as staining for known populations markers. Immune cells, CD45; fibroblast, CD34; endothelial, CD31; mTEC, UEA1; cTEC, Ly51; mature mTEC, MHC-II. The red border marks stroma single-cell sorting gate. **b**, Index sort tracks showing the intensity of protein levels for MHC-II, UEA1 and Ly51 in individual single cells

shown in Fig. 1a–c. **c**, Cell–cell correlation of CD45$^-$ thymic stroma cells calculated over 132 differentially expressed genes. **d**, Pairwise distance distribution between cells within the three main stromal lineages. Distance is defined as 1 − Spearman. Box plots display median bar, first–third quantile box and 5th–95th percentile whiskers. $n = 1{,}825$ single cells. **e**, **f**, Gene expression profiles of 749 cells from fibroblast clusters of the CD45$^-$ stroma, marked by increased expression of *Col1a1* (**e**, Supplementary Table 1) and 221 cells from the endothelial clusters of the CD45$^-$ stroma, marked by increased expression of *Pecam1* (**f**, Supplementary Table 1).

AQ6

Extended Data Fig. 3

Thymic epithelial cells are characterized by four subsets of mTEC and a single cTEC subset.

a, Heat map showing a metacell analysis of 2,341 thymic epithelial cells (CD45$^-$EpCAM$^+$), featuring the 73 most variable genes, from 15 biological replicates of 4–6-week-old mice. Colour bar represents separation of 36 metacells into five main populations. **b**, Two-dimensional graph representation of the metacell model in Fig. 1a (see Methods). Big circles represent metacells, and are colour-coded as shown in **a**. **c**, FACS index sorting measurement of Ly51 and UEA1 in epithelial cells. Cells are coloured based on cluster association as determined in **a**. Dashed lines outline Ly51$^+$UEA1$^-$ and Ly51$^-$UEA1$^+$ gates. **d**, Fraction of TEC subsets out of Ly51$^+$UEA1$^-$ and Ly51$^-$UEA1$^+$ populations, assessed by gating single cells on index sorting protein measurements of UEA1 and Ly51 in **c**. **e**, Controlling for batch effect as determined by the relative share of each batch in all metacells. Batches are ordered by biological replicate (marked by dashed lines) and sorting scheme (either CD45$^-$ or CD45$^-$EpCAM$^+$). **f**, **g**, Single molecule FISH assay on 5–8-μm-thick cryosections (see Methods) using fluorescent probes against the genes *Epcam* and *Sbsn* (**f**) or *Avil* (**g**). Blue, DAPI. The experiments were repeated independently four times with similar results. **h**, Immunofluorescence images of the protein markers: Pigr and Dclk1. Medulla (M) and cortex (C) are separated by dashed lines, distinguished by nuclei density. Blue, DAPI. **f**–**h**, Scale bars, 20 μm. The experiments were repeated independently twice with similar results. **i**, Projection of representative differentially expressed genes onto the two-dimensional graph of epithelial cells.

Extended Data Fig. 4

TEC dynamics during thymus development.

a, Flow cytometry scheme of thymic epithelial cells from different development time points. Numbers indicate fraction of $CD45^{-}EpCAM^{+}$ cells. **b**, Summary of the *K* (*K* = 50) nearest neighbours of embryonic cells, grouped into metacells. Neighbours of adult origin are coloured by TEC subsets, and of embryonic origin are coloured by developmental time point. Metacells with more than 20% adult neighbours were assigned to a TEC subset. **c**, Normalized mean expression of differential genes across TEC mature populations (4 weeks old) and unassigned cells from developmental time points (grey). $n = 4$ (E14.5), 3 (E18.5) and 2 (6 days) independent animals. Data are median (bars) and individual animals (dots). **d**, Differential gene expression between early cTEC (e-cTEC) from three developmental time points and the mature cTEC. Axes represent UMI count per 1,000 UMI, normalized to cell numbers. **e**, Distribution of cell cycle gene expression across cells from developing TEC is bimodal. The red line indicates the empirical proliferation threshold. **f**, Frequency of proliferating cells in the epithelial population at each developmental time point. Colour code as in **b**. **g**, Gene pairwise Spearman correlation over 2,319 e-cTEC single cells reveals three gene modules jointly expressed across embryonic early TEC and mature cTEC populations. **h**, GO annotations enrichment analysis of the three cTEC gene modules. For cell cycle, $n = 114$; antigen presentation, $n = 33$; and antigen processing, $n = 56$.

Extended Data Fig. 5

Genetic and epigenetic characterization of TEC subsets.

a, In silico gating of mTEC I–IV populations by index sorting measurements of surface markers. The same gating schemes were used to purify these populations by FACS (Fig. 2a). Cells are colour-coded as shown in Fig. 1. Blue, cTEC; light blue, mTEC I; red, mTEC II; yellow, mTEC III; green, mTEC IV. **b**, Relative enrichment (log_2 fold change compared to total $Cd45^{-}EpCAM^{+}$ epithelial cells) of the individual mTEC I–IV subsets gated according to **a**. **c**, Heat map showing pairwise Spearman

correlation of 29,472 H3K4me2 peaks (top) or ATAC-seq peaks (bottom) from mTEC I–IV sorted populations. Biological replicates for each population are shown. **d**, Scatter plots of mTEC I–IV H3K4me2 ChIP–seq peaks in biological duplicates. **e**, Summary of the most significant motifs enriched in each cluster of mTEC I–IV H3K4me2 differential peaks (Fig. 2e). *P* values are derived from binomial tests after FDR correction for multiple hypotheses. $n = 2{,}302$ differential peaks.

AQ7

Extended Data Fig. 6

Characterization of AIRE-dependent mTEC subsets.

a, Variance of genes plotted against their mean value (genes with >50 total UMI are shown). Orange dots indicate variable genes. **b**, Pairwise Pearson gene correlations in AIRE-dependent (left) and AIRE-independent (right) TRA gene lists across 3,074 TEC single cells. Levels of differential expression (highest change of expression in cluster compared to median across all clusters) are indicated as bars; bar colours indicate cluster association to TEC population. **c**, Comparison of stochastic gene expression between *Aire* knockout and wild-type cells in mTEC III and IV populations. Marginal distribution is shown as histogram. Axes represent UMI count per 1,000 UMI, normalized to cell numbers. **d**, Flow cytometry scheme of thymic *Aire* knockout cells showing the percentage of each TEC population compared to wild-type percentage (shown in brackets). **e**, Representative immunofluorescence images of two independent experiments, for tdTomato across different organs. In the thymus, the medulla (M) and cortex (C) are separated by dashed lines, distinguished by nuclei density. Blue, DAPI. Scale bars, 100 μm. **f**, qPCR analysis of *Csnb* (left) and *cre* (right) genes in $Csnb^{cre+}Rosa26^{tdTomato}$ (cre^{+}) and $Csnb^{cre-}Rosa26^{tdTomato}$ (cre^{-}) across thymic populations. Dot plots display mean and error bars indicate s.e.m. $n = 2$ (wild type) or $n = 3$ ($Csnb^{cre+}$) biologically independent animals. **g**, Flow cytometric analysis of tdTomato expression in mTEC subsets (colours as in Fig. 1) isolated from thymi of $Csnb^{cre+}Rosa26^{tdTomato}$- or $Csnb^{cre-}Rosa26^{tdTomato}$-reporter mice.

Extended Data Fig. 7

Comparing intestinal tuft cells and the mTEC IV population.

a, Flow cytometry scheme of mTEC IV cells (*Sox9*-eGFP^{+}L1CAM^{+}) sorting. **b**, Flow cytometry scheme of small intestine *Hpgds*-dTomato^{+} tuft cell sorting. **c**, Heat map showing gene expression profiles across 1,903 intestinal tuft (*Hpgds*-tdTomato^{+}) single cells, grouped into 68 metacells. **d**, Comparison of gene expression between tuft cells isolated from small intestine (*Hpgds*-tdTomato^{+}; *x* axis) and mTEC IV cells isolated from thymus (CD45^{-}EpCAM^{+}*Sox9*–eGFP^{+}L1CAM^{+}; *y* axis). Axes represent UMI count per 1,000 UMI, normalized to cell numbers. **e**, Normalized mean expression of differential genes across TEC populations, sorted mTEC IV cells (L1CAM^{+}*Sox9*-eGFP^{+}) and intestinal Tuft (*Hpgds*-tdTomato^{+}) cells. **f**, GO annotations enrichment in differential genes (fold change >2) between intestinal tufts ($n = 634$) and mTEC IV cells (L1CAM^{+}*Sox9*-eGFP^{+}) ($n = 1{,}308$).

Extended Data Fig. 8

The transcription factor Pou2f3 is a master regulator of mTEC IV.

a, Flow cytometry scheme for sorting of EpCAM^{+} cells from *Pou2f3* knockout thymi. **b**, Projection of representative TEC subtype-specific markers onto the two-dimensional mapping of *Pou2f3* wild-type and knockout cells to the epithelial model of Fig. 1 (see Methods). **c**, Bar plot showing $\log_2$ fold change between TEC subpopulation abundances in *Pou2f3* knockout ($n = 451$ single cells) and wild-type ($n = 1121$) mice. Error bars represent 95% confidence intervals. **d**, Pooled expression of mTEC IV genes across cells from *Pou2f3* knockout and wild-type (WT) mice. The *x* axis represents UMI count per 1,000 UMI; *y* axis represents fraction of expression from total UMI count of the cells. Green cells indicate cells classified as mTEC IV (two-sided binomial test; FDR-adjusted $P < 10^{-10}$). $n = 1{,}572$ single cells. **e**, Differential gene expression between mTEC I–III cells isolated from control (wild-type) and *Pou2f3* knockout mice. Axes represent UMI count per 1,000 UMI, normalized to cell numbers.

Extended Data Fig. 9

mTEC IV shape the thymus immune niche.

a, Heat map showing a metacell analysis of 3,500 $CD45^+IL\text{-}25R^+$ cells across five clusters. **b**, Differential gene expression in ILC2 cells from two biological replicates compared to other $CD45^+IL\text{-}25R^+$ cells (rest) from each replicate. Axes represent $\log_2$ fold change. **c**, Percentages of $CD45^+$ cells in *Pou2f3* knockout and wild-type thymi, determined by flow cytometry. Circles and diamonds indicate independent mice, centre line indicates the mean value. **d**, Flow cytometry analysis of CD4-, CD8-, CD25- and CD44-expressing cells in *Pou2f3* knockout and wild-type thymi. The experiment was repeated independently four times with similar results to confirm reproducibility. **e**, Flow cytometry sorting scheme of thymic $CD45^+IL\text{-}25R^+$ cells from *Pou2f3* knockout and wild-type thymi. **f**, Percentage of $CD45^+IL\text{-}25R^+$ cells in *Pou2f3* knockout and wild-type thymi. Circles and diamonds indicate independent mice, centre line indicates the mean value. **g**, Percentages (left) and numbers (right) of ILC2 ($Lin^-TCR^-CD127^+GATA3^+Ror\gamma t^-$) cells within the single-cell gate in *Pou2f3* knockout and wild-type mice. Circles and diamonds indicate independent mice, centre line indicates the mean value. A one-tailed Student's *t*-test was used for the comparison, $*P < 0.05$.

Extended Data Table 1

Regions of enhancer enrichment peaks

Regions of enhancer enrichment peaks.

a, H3K4me2 iChIP–seq peak calling. **b**, ATAC peak calling.

Supplementary information

Supplementary Information

This file contains Supplementary Note 1: MetaCell - Correcting and clustering single cell RNA-seq data using k-nn graph covering.

Reporting Summary

Supplementary Table 1

Average gene expression across CD45- thymic fibroblasts and endothelial cells.

Supplementary Table 2

Average gene expression across CD45- thymic epithelial cells.

Supplementary Table 3

Average gene expression across the development of thymic epithelial cells (E14.5, E18.5 & 6 days PN).

Supplementary Table 4

Definition of stochastic genes and TRA.

References

1. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat. Rev. Immunol*. **14**, 377–391 (2014).

2. Abramson, J. & Anderson, G. Thymic epithelial cells. *Annu. Rev. Immunol*. **35**, 85–118 (2017).

3. Takahama, Y., Ohigashi, I., Baik, S. & Anderson, G. Generation of diversity in thymic epithelial cells. *Nat. Rev. Immunol*. **17**, 295–305 (2017).

4. Jaitin, D. A. et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).

5. Paul, F. et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* **163**, 1663–1677 (2015).

6. Takada, K. & Takahama, Y. Positive-selection-inducing self-peptides displayed by cortical thymic epithelial cells. *Adv. Immunol*. **125**, 87–110 (2015).

7. Wong, K. et al. Multilineage potential and self-renewal define an epithelial progenitor cell population in the adult thymus. *Cell Rep*. **8**, 1198–1209 (2014).

8. Galliano, M. F. et al. Characterization and expression analysis of the *Spink5* gene, the mouse ortholog of the defective gene in Netherton syndrome. *Genomics* **85**, 483–492 (2005).

9. Hale, L. P. & Markert, M. L. Corticosteroids regulate epithelial cell differentiation and Hassall body formation in the human thymus. *J. Immunol*. **172**, 617–624 (2004).

10. Lara-Astiaso, D. et al. Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).

11. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

12. Brennecke, P. et al. Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. *Nat. Immunol*. **16**, 933–941 (2015).

13. Meredith, M., Zemmour, D., Mathis, D. & Benoist, C. Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat. Immunol*. **16**, 942–949 (2015).

14. Gerbe, F. et al. Distinct ATOH1 and Neurog3 requirements define tuft cells as a new secretory cell type in the intestinal epithelium. *J. Cell Biol*. **192**, 767–780 (2011).

15. Gerbe, F. et al. Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites. *Nature* **529**, 226–230 (2016).

16. Gerbe, F., Brulin, B., Makrini, L., Legraverend, C. & Jay, P. DCAMKL-1 expression identifies tuft cells rather than stem cells in the adult mouse intestinal epithelium. *Gastroenterology*

AQ8

137, 2179–2180 (2009).

17. Tanimizu, N., Nishikawa, Y., Ichinohe, N., Akiyama, H. & Mitaka, T. Sry HMG box protein 9-positive ($Sox9^{+}$) epithelial cell adhesion molecule-negative (EpCAM−) biphenotypic cells derived from hepatocytes are involved in mouse liver regeneration. *J. Biol. Chem*. **289**, 7589–7598 (2014).

18. Gury-BenAri, M. et al. The spectrum and regulatory landscape of intestinal innate lymphoid cells are shaped by the microbiome. *Cell* **166**, 1231–1246 (2016).

19. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).

20. Scialdone, A. et al. Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**, 54–61 (2015).

21. Sansom, S. N. et al. Population and single-cell genomics reveal the *Aire* dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome Res*. **24**, 1918–1931 (2014).

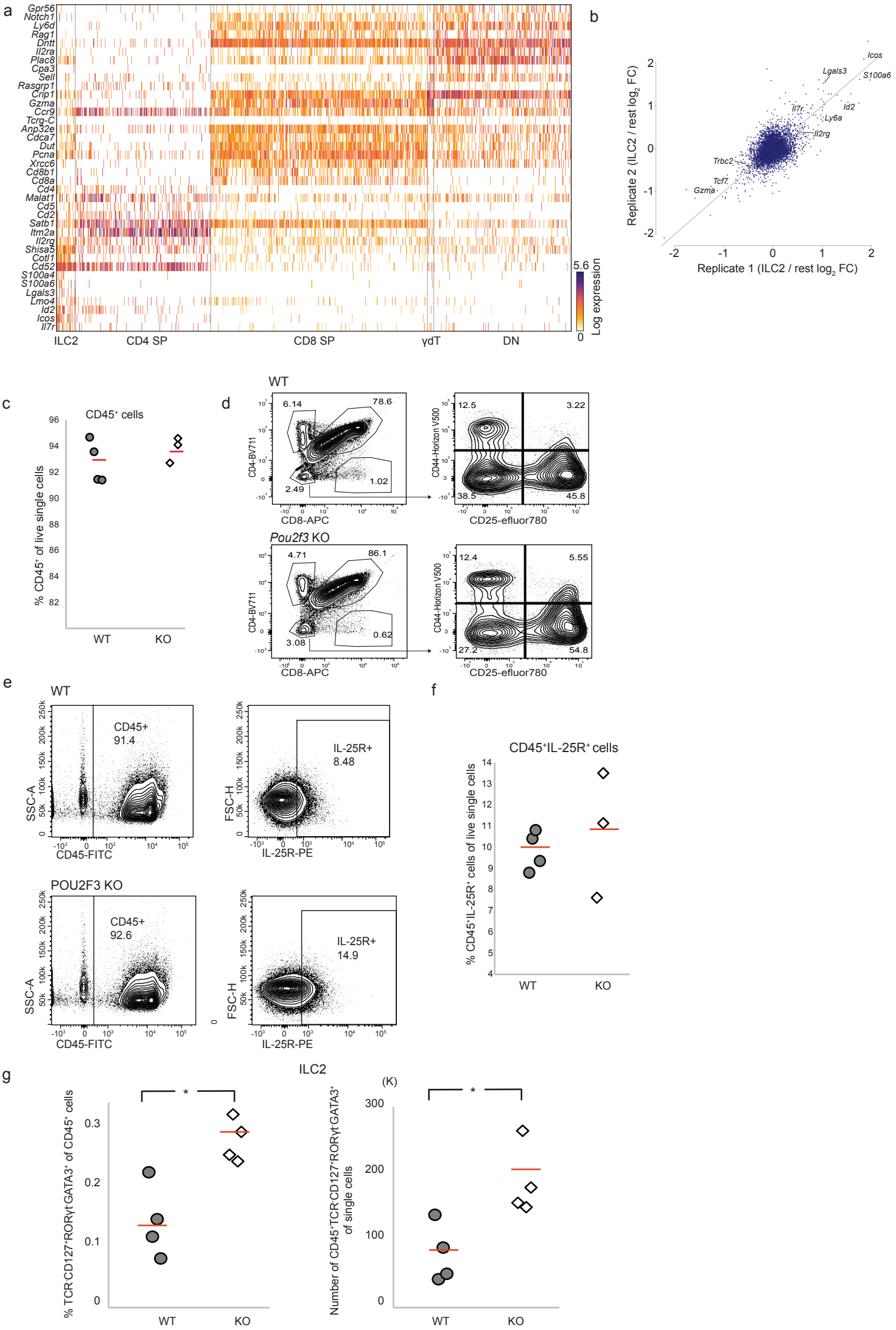
22. Lyubimova, A. et al. Single-molecule mRNA detection and counting in mammalian tissue. *Nat. Protoc*. **8**, 1743–1758 (2013).

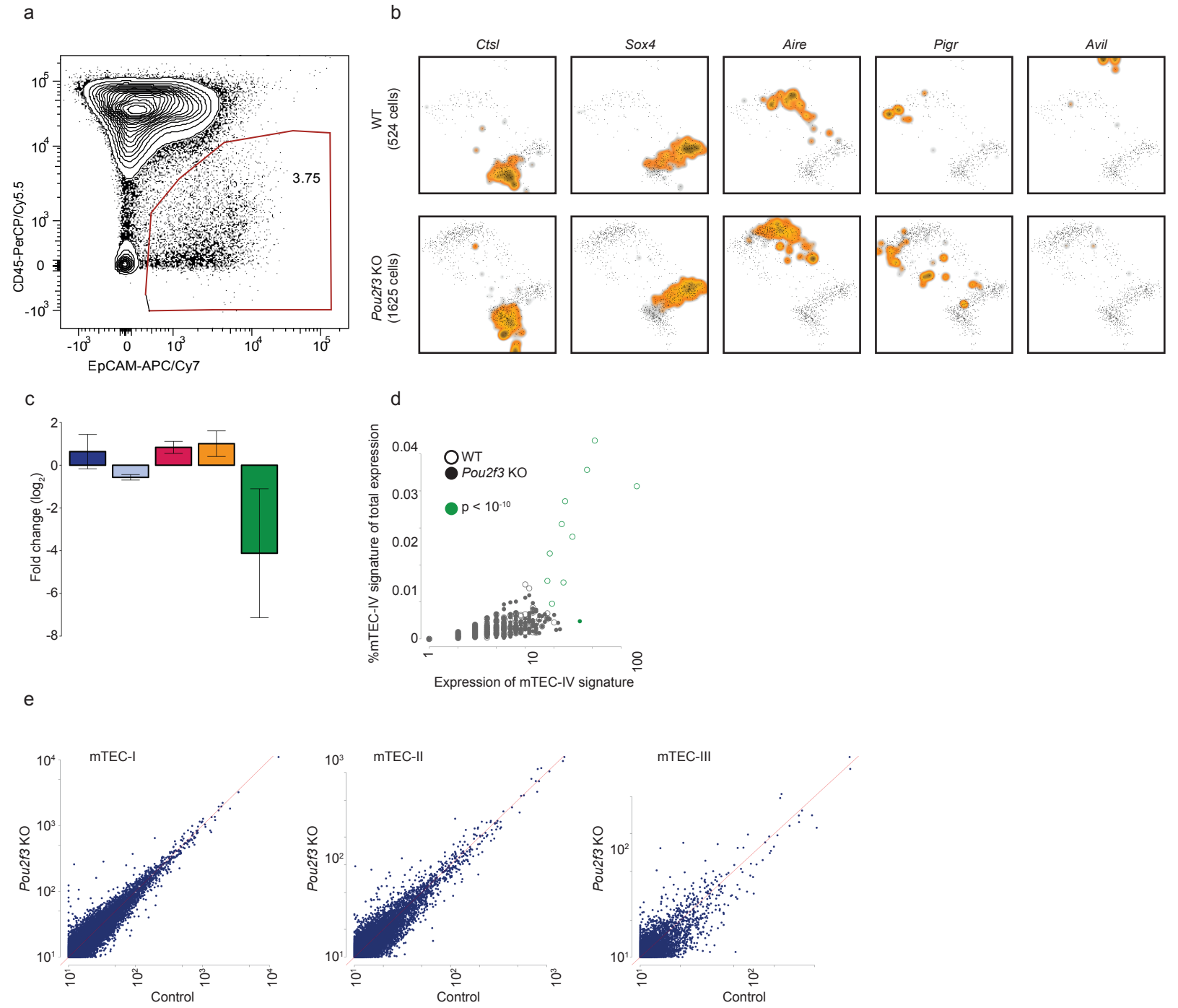
23. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol*. **109**, 21.29.1–21.29.9 (2015).

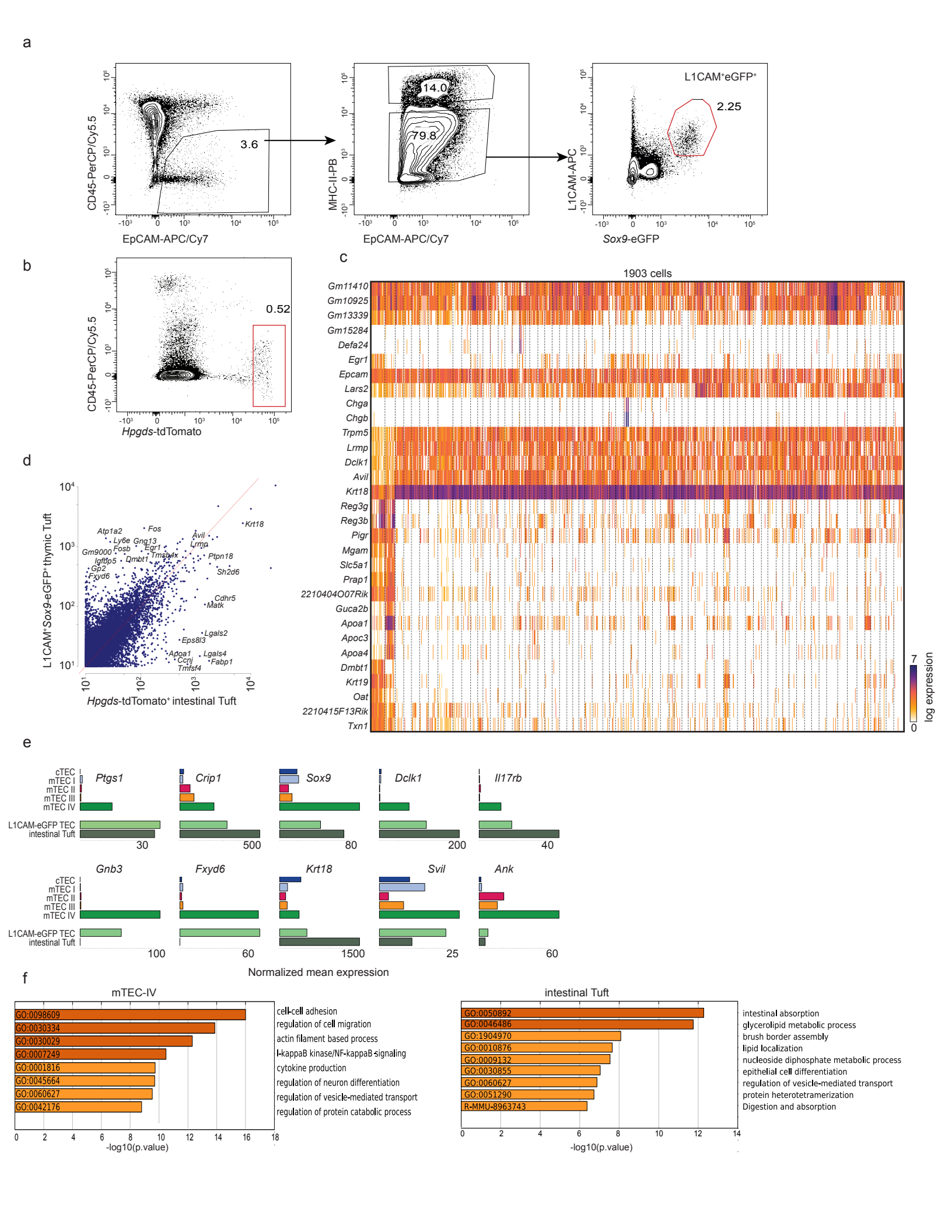
24. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. **10**, R25 (2009).

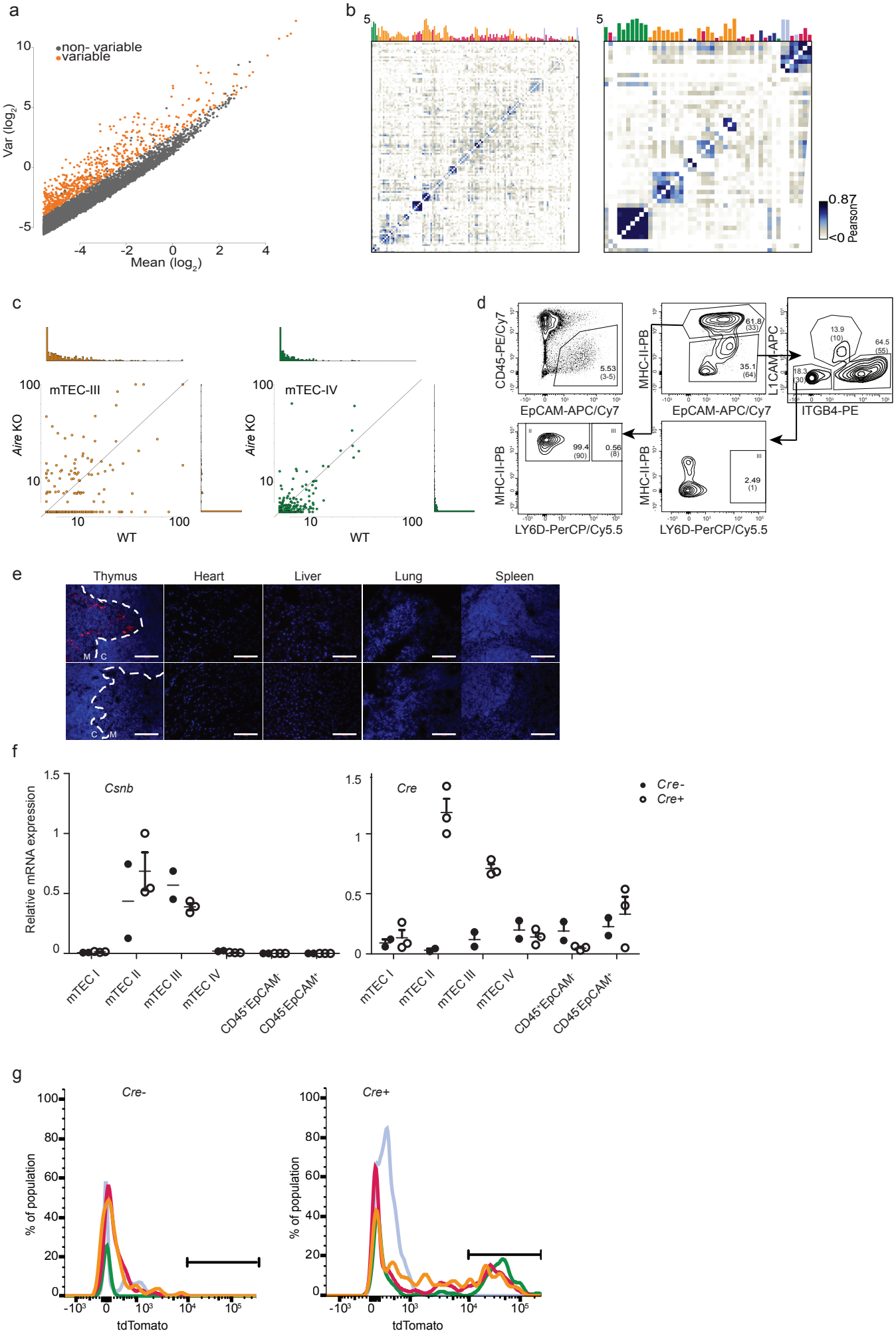
25. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

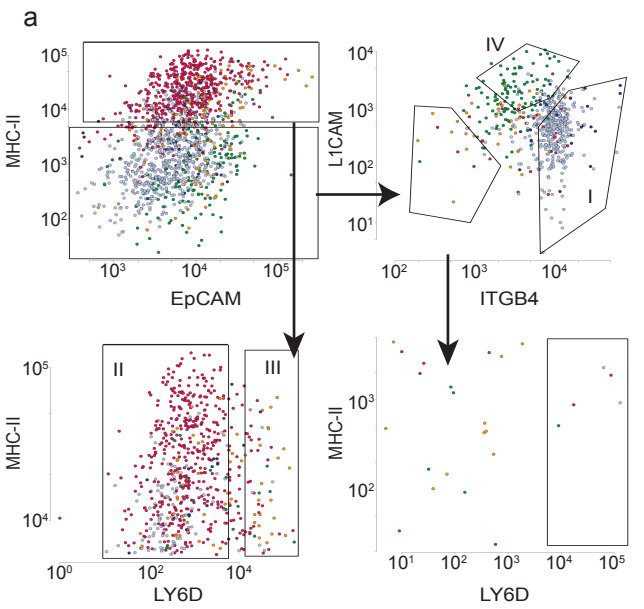
26. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat*. **5**, 1752–1779 (2011).

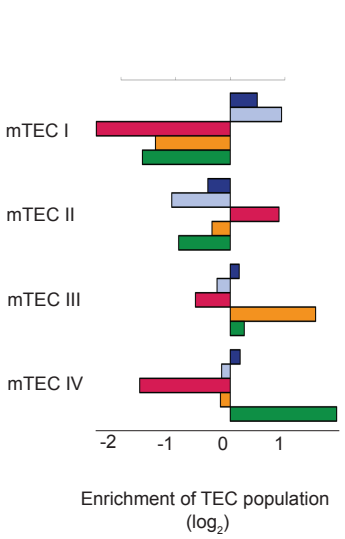

a
Gpr56
Notch1
Ly6d
Rag1
Dntt
Il2ra
Plac8
Cpa3
Sell
Rasgrp1
Crip1
Gzma
Ccr9
Tcrg-C
Anp32e
Cdca7
Dut
Pcna
Xrcc6
Cd8b1
Cd8a
Cd4
Malat1
Cd5
Cd2
Satb1
Itm2a
Il2rg
Shisa5
Cotl1
Cd52
S100a4
S100a6
Lgals3
Lmo4
Id2
Icos
Il7r
ILC2
CD4 SP
CD8 SP
γdT
DN
5.6
0
Log expression
b
Replicate 2 (ILC2 / rest log2 FC)
Replicate 1 (ILC2 / rest log2 FC)
Icos
Lgals3
S100a6
Il7r
Id2
Ly6a
Il2rg
Trbc2
Tcf7
Gzma
c
CD45+ cells
% CD45+ of live single cells
WT
KO
d
WT
6.14
78.6
2.49
1.02
CD8-APC
CD4-BV711
12.5
3.22
38.5
45.8
CD25-efluor780
CD44-Horizon V500
Pou2f3 KO
4.71
86.1
3.08
0.62
12.4
5.55
27.2
54.8
e
WT
CD45+
91.4
SSC-A
CD45-FITC
IL-25R+
8.48
FSC-H
IL-25R-PE
POU2F3 KO
CD45+
92.6
IL-25R+
14.9
f
CD45+IL-25R+ cells
% CD45+IL-25R+ cells of live single cells
WT
KO
g
ILC2
% TCR-CD127+RORγt+GATA3+ of CD45+ cells
Number of CD45+TCR-CD127+RORγt+GATA3+ of single cells
(K)
WT
KO
*

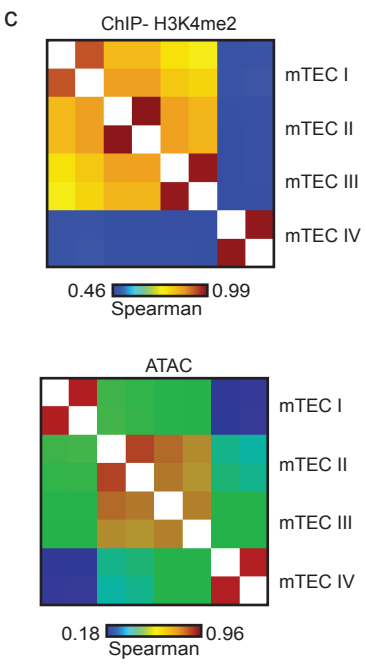

a
b
c
d
e
CD45-PerCP/Cy5.5
EpCAM-APC/Cy7
3.75
Ctsl
Sox4
Aire
Pigr
Avil
WT (524 cells)
Pou2f3 KO (1625 cells)
Fold change (log2)
%mTEC-IV signature of total expression
Expression of mTEC-IV signature
WT
Pou2f3 KO
p < 10-10
mTEC-I
mTEC-II
mTEC-III
Pou2f3 KO
Control

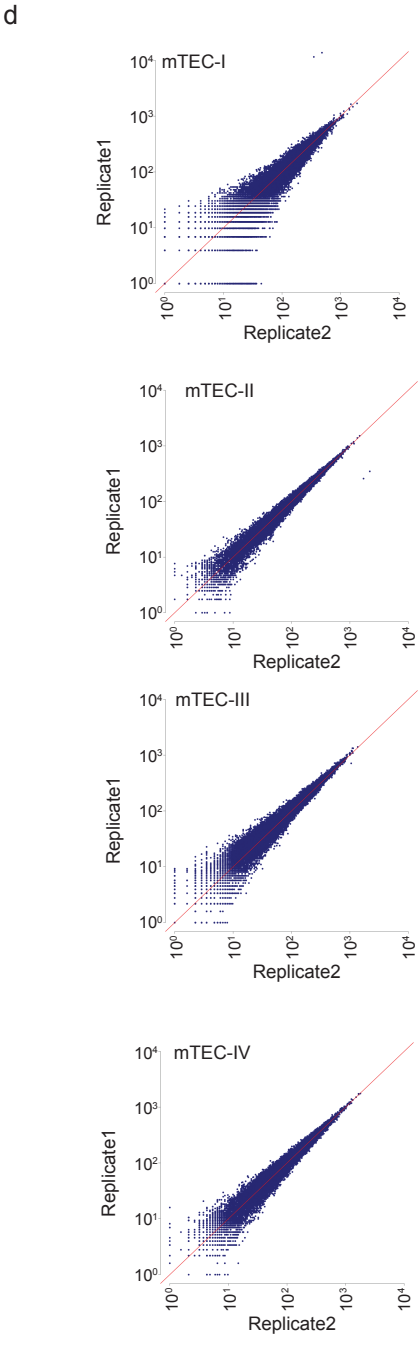

a
CD45-PerCP/Cy5.5
EpCAM-APC/Cy7
3.6
MHC-II-PB
14.0
79.8
L1CAM-APC
Sox9-eGFP
L1CAM+eGFP+
2.25
b
Hpgds-tdTomato
0.52
c
1903 cells
Gm11410
Gm10925
Gm13339
Gm15284
Defa24
Egr1
Epcam
Lars2
Chga
Chgb
Trpm5
Lrmp
Dclk1
Avil
Krt18
Reg3g
Reg3b
Pigr
Mgam
Slc5a1
Prap1
2210404O07Rik
Guca2b
Apoa1
Apoc3
Apoa4
Dmbt1
Krt19
Oat
2210415F13Rik
Txn1
log expression
d
L1CAM+Sox9-eGFP+ thymic Tuft
Hpgds-tdTomato+ intestinal Tuft
e
Ptgs1
Crip1
Sox9
Dclk1
Il17rb
Gnb3
Fxyd6
Krt18
Svil
Ank
cTEC
mTEC I
mTEC II
mTEC III
mTEC IV
L1CAM-eGFP TEC
intestinal Tuft
Normalized mean expression
f
mTEC-IV
GO:0098609
GO:0030334
GO:0030029
GO:0007249
GO:0001816
GO:0045664
GO:0060627
GO:0042176
cell-cell adhesion
regulation of cell migration
actin filament based process
I-kappaB kinase/NF-kappaB signaling
cytokine production
regulation of neuron differentiation
regulation of vesicle-mediated transport
regulation of protein catabolic process
-log10(p.value)
intestinal Tuft
GO:0050892
GO:0046486
GO:1904970
GO:0010876
GO:0009132
GO:0030855
GO:0060627
GO:0051290
R-MMU-8963743
intestinal absorption
glycerolipid metabolic process
brush border assembly
lipid localization
nucleoside diphosphate metabolic process
epithelial cell differentiation
regulation of vesicle-mediated transport
protein heterotetramerization
Digestion and absorption
-log10(p.value)


a
non- variable
variable
Var (log2)
Mean (log2)
b
0.87
<0
Pearson
c
mTEC-III
mTEC-IV
Aire KO
WT
d
CD45-PE/Cy7
EpCAM-APC/Cy7
MHC-II-PB
LY6D-PerCP/Cy5.5
L1CAM-APC
ITGB4-PE
e
Thymus
Heart
Liver
Lung
Spleen
f
Csnb
Cre
Cre-
Cre+
Relative mRNA expression
mTEC I
mTEC II
mTEC III
mTEC IV
CD45+EpCAM-
CD45-EpCAM+
g
Cre-
Cre+
% of population
tdTomato



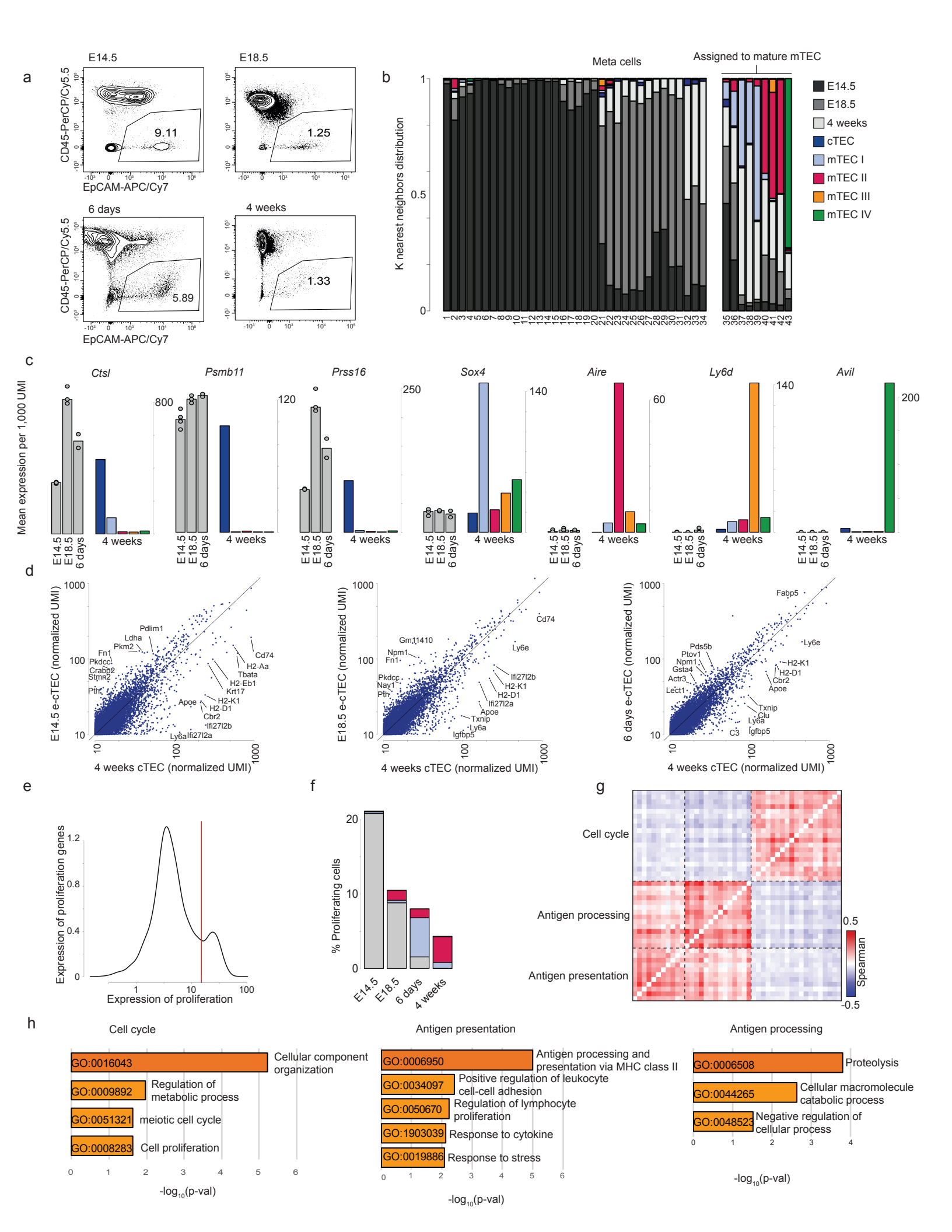


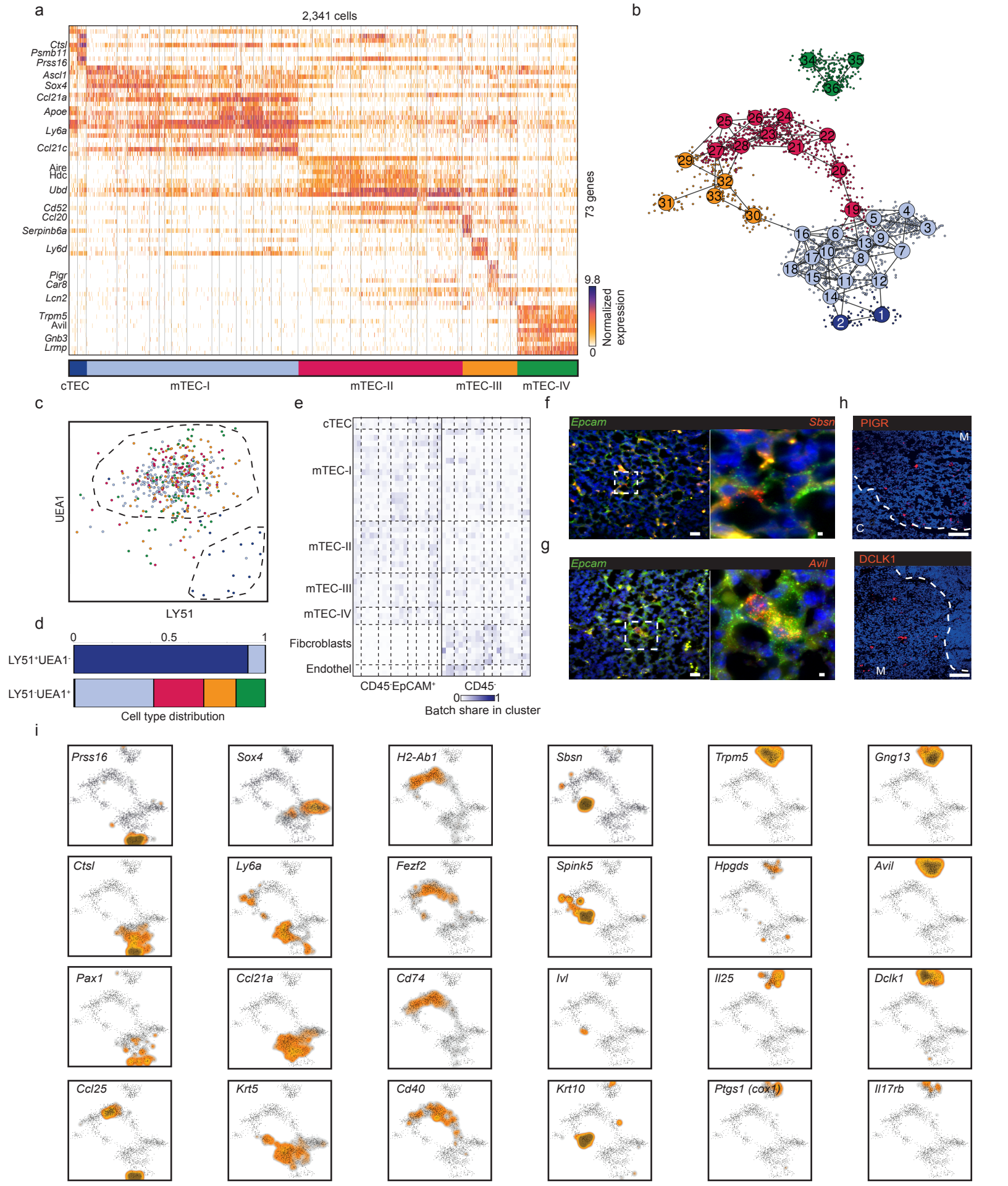


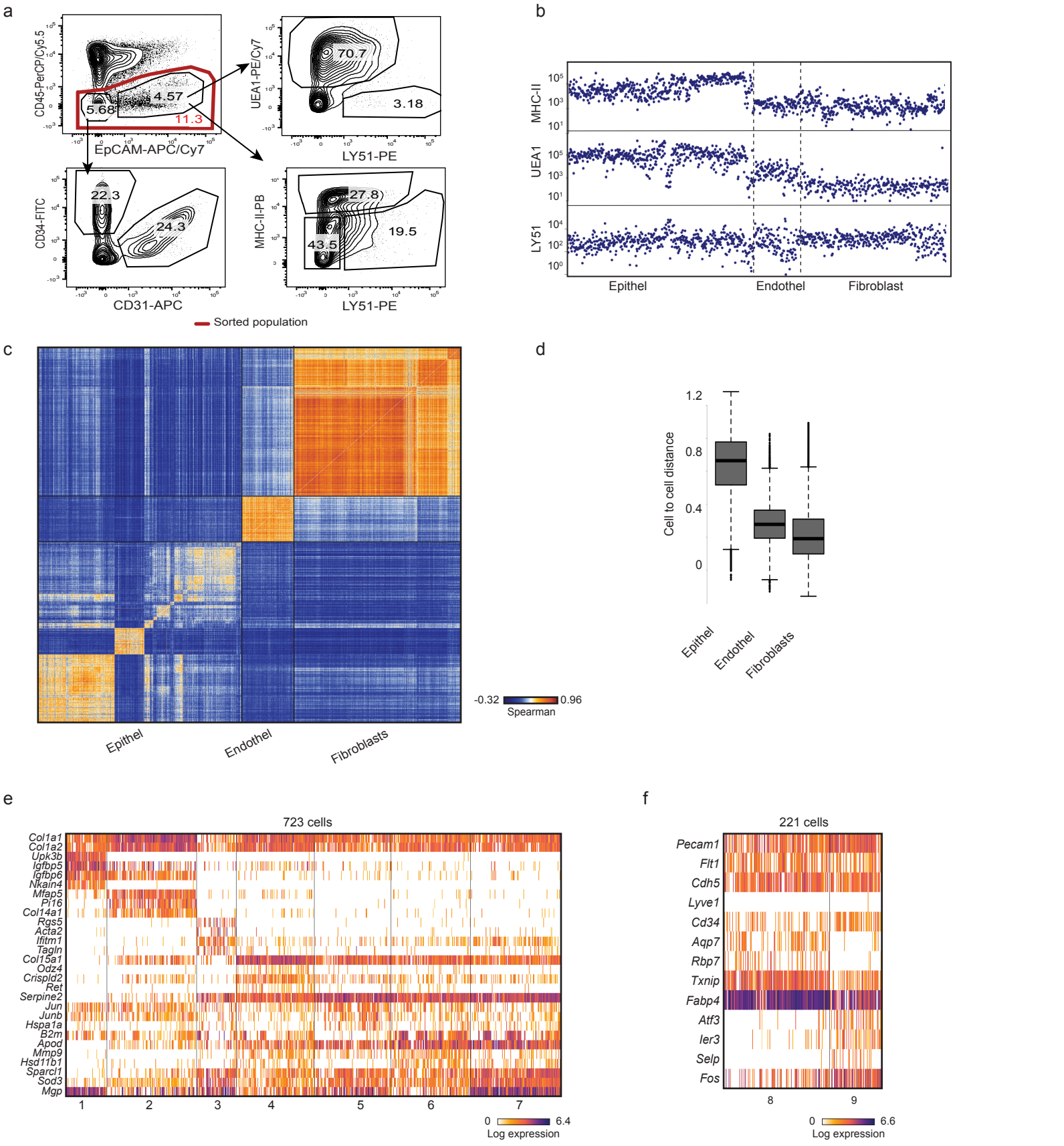


e

cluster	rank	pval	% found	Fold change	motif	
1	1	10^{-33}	0.26	21.09	MF0003.1_REL_class	
2	1	10^{-109}	0.82	12.95	MF0003.1_REL_class	
3	1	10^{-46}	0.31	17.10	MF0003.1_REL_class	
	2	10^{-14}	0.05	122.75	MZF1(var.2)/MA0057.1	
4	1	10^{-17}	0.30	5.18	NFkB-p65(RHD)/GM12787p65	
	2	10^{-15}	0.30	4.35	AP-1(bZIP)/ThioMacPU.1	
5	1	10^{-27}	0.39	6.49	POU2F2/MA0507.1	
	2	10^{-16}	0.21	8.02	ELF5(ETS)/T47DELF5	
	3	10^{-15}	0.09	37.20	HNF1A/MA0046.2	
6	1	10^{-41}	0.33	11.31	POU2F2/MA0507.1	
	2	10^{-20}	0.07	130.20	HNF1A/MA0046.2	
	3	10^{-14}	0.07	33.81	NFkB-p65(RHD)/GM12787p65	
7	1	10^{-805}	0.69	9.09	POU2F2/MA0507.1	
	2	10^{-28}	0.08	3.21	Fox:Ebox(Forkhead,bHLH)/Panc1-Foxa2	
	3	10^{-25}	0.18	1.94	Ets1-distal(ETS)/CD4+PolII	


a
E14.5
E18.5
6 days
4 weeks
CD45-PerCP/Cy5.5
EpCAM-APC/Cy7
9.11
1.25
5.89
1.33
b
Meta cells
Assigned to mature mTEC
K nearest neighbors distribution
E14.5
E18.5
4 weeks
cTEC
mTEC I
mTEC II
mTEC III
mTEC IV
c
Mean expression per 1,000 UMI
Ctsl
Psmb11
Prss16
Sox4
Aire
Ly6d
Avil
E14.5
E18.5
6 days
4 weeks
d
E14.5 e-cTEC (normalized UMI)
E18.5 e-cTEC (normalized UMI)
6 days e-cTEC (normalized UMI)
4 weeks cTEC (normalized UMI)
e
Expression of proliferation genes
Expression of proliferation
f
% Proliferating cells
g
Cell cycle
Antigen processing
Antigen presentation
Spearman
h
Cell cycle
GO:0016043 Cellular component organization
GO:0009892 Regulation of metabolic process
GO:0051321 meiotic cell cycle
GO:0008283 Cell proliferation
Antigen presentation
GO:0006950 Antigen processing and presentation via MHC class II
GO:0034097 Positive regulation of leukocyte cell-cell adhesion
GO:0050670 Regulation of lymphocyte proliferation
GO:1903039 Response to cytokine
GO:0019886 Response to stress
Antigen processing
GO:0006508 Proteolysis
GO:0044265 Cellular macromolecule catabolic process
GO:0048523 Negative regulation of cellular process
-log10(p-val)


a
2,341 cells
Ctsl
Psmb11
Prss16
Ascl1
Sox4
Ccl21a
Apoe
Ly6a
Ccl21c
Aire
Hdc
Ubd
Cd52
Ccl20
Serpinb6a
Ly6d
Pigr
Car8
Lcn2
Trpm5
Avil
Gnb3
Lrmp
73 genes
9.8
Normalized expression
0
cTEC
mTEC-I
mTEC-II
mTEC-III
mTEC-IV
b
c
UEA1
LY51
d
0
0.5
1
LY51+UEA1-
LY51-UEA1+
Cell type distribution
e
cTEC
mTEC-I
mTEC-II
mTEC-III
mTEC-IV
Fibcroblasts
Endothel
CD45-EpCAM+
CD45-
0
1
Batch share in cluster
f
Epcam
Sbsn
g
Epcam
Avil
h
PIGR
M
C
DCLK1
M
i
Prss16
Sox4
H2-Ab1
Sbsn
Trpm5
Gng13
Ctsl
Ly6a
Fezf2
Spink5
Hpgds
Avil
Pax1
Ccl21a
Cd74
Ivl
Il25
Dclk1
Ccl25
Krt5
Cd40
Krt10
Ptgs1 (cox1)
Il17rb


a
CD45-PerCP/Cy5.5
EpCAM-APC/Cy7
5.68
4.57
11.3
UEA1-PE/Cy7
70.7
3.18
LY51-PE
CD34-FITC
CD31-APC
22.3
24.3
MHC-II-PB
LY51-PE
27.8
19.5
43.5
Sorted population
b
MHC-II
UEA1
LY51
Epithel
Endothel
Fibroblast
c
Epithel
Endothel
Fibroblasts
-0.32
0.96
Spearman
d
Cell to cell distance
1.2
0.8
0.4
0
Epithel
Endothel
Fibroblasts
e
723 cells
Col1a1
Col1a2
Upk3b
Igfbp5
Igfbp6
Nkain4
Mfap5
Pi16
Col14a1
Rgs5
Acta2
Ifitm1
Tagln
Col15a1
Odz4
Crispld2
Ret
Serpine2
Jun
Junb
Hspa1a
B2m
Apod
Mmp9
Hsd11b1
Sparcl1
Sod3
Mgp
1
2
3
4
5
6
7
0
6.4
Log expression
f
221 cells
Pecam1
Flt1
Cdh5
Lyve1
Cd34
Aqp7
Rbp7
Txnip
Fabp4
Atf3
Ier3
Selp
Fos
8
9
0
6.6
Log expression

a

Organ	Gating	Age	Strain		Mouse	nbatches	ncells
Thymus	CD45$^-$	4-6 weeks	WT	a	1	3	322
					2	3	358
					3	4	393
					4	2	206
					5	2	188
					6	5	332
					7	2	222
	CD45$^-$ EpCAM$^+$	4-6 weeks	WT	b	8	2	208
					9	4	267
					10	3	168
					11	4	443
					12	2	95
					13	3	178
					14	2	246
					15	1	111
			Aire KO	c	16	11	1332
		E14.5	WT	d	17	2	269
					18	2	274
					19	4	525
					20	2	275
		E18.4	WT	e	21	1	129
					22	2	220
					23	5	546
		6 days	WT	f	24	4	396
					25	4	440
		4-6 weeks	*Pou2f3* Control	g	26	4	339
					27	1	112
			Pou2f3 KO	h	28	11	549
					29	6	572
	CD45$^-$ EpCAM$^+$ *Sox9*-eGFP$^+$	4-6 weeks	*Sox9*^{eGFP}	i	30	2	376
Intestine	EpCAM$^+$ *Hpgds*-tdTomato$^+$	4-6 weeks	*Hpgds*^{tdTomato}	j	31	4	928
					32	4	951
Thymus	CD45$^+$ IL-25R$^+$	4-6 weeks	WT	k	33	4	1075
					34	6	1560
					35	4	865
			Pou2f3 Control	l	36	2	613
					37	2	217
					38	2	541
			Pou2f3 KO	m	39	2	480
					40	2	333
					Total	135	17654