European Journal of Immunology

Supporting Information

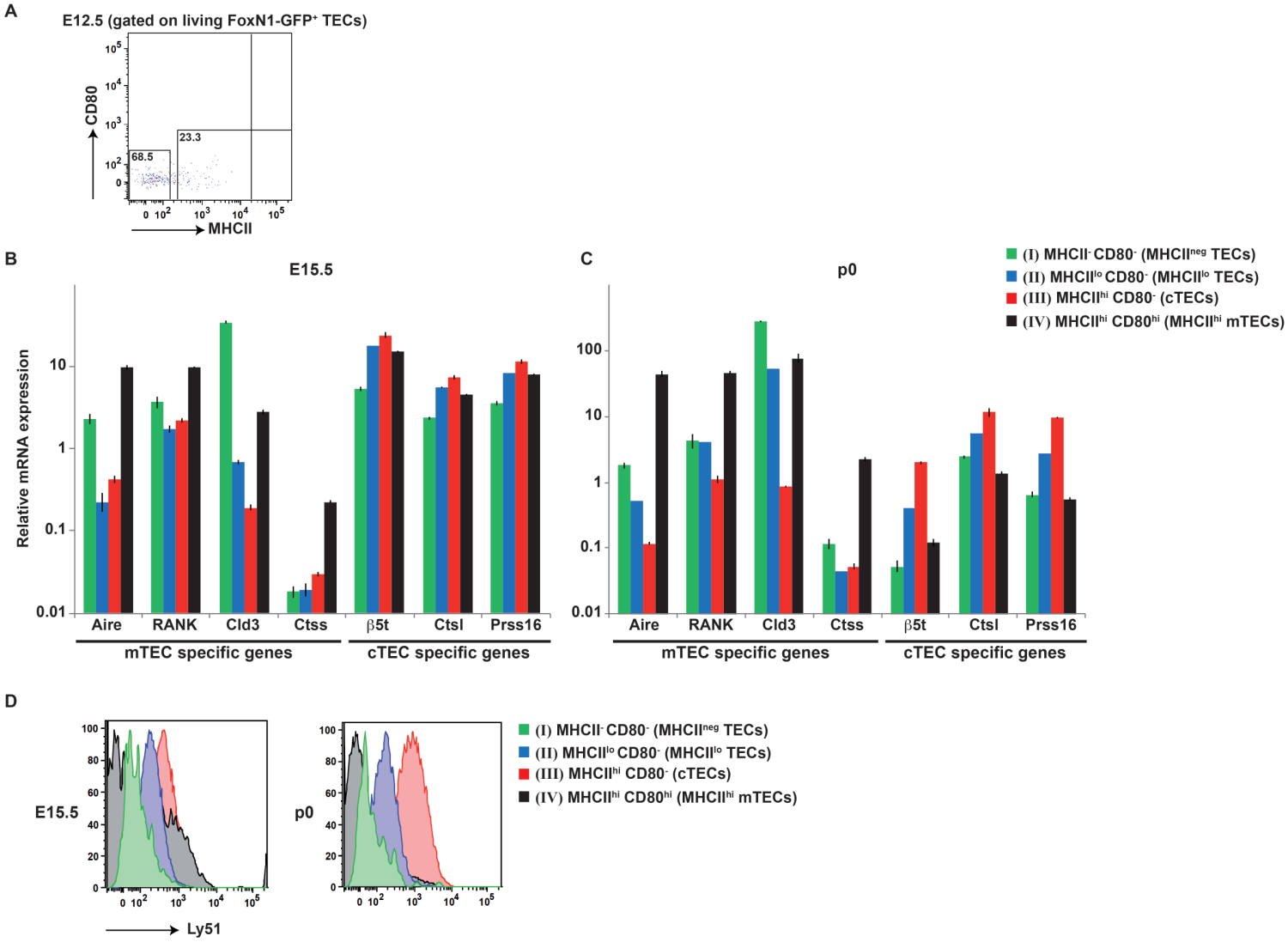
for

DOI 10.1002/eji.201747006

Fabian Brunk, Chloé Michel, Tim Holland-Letz, Alla Slynko, Annette Kopp-Schneider, Bruno Kyewski and Sheena Pinto

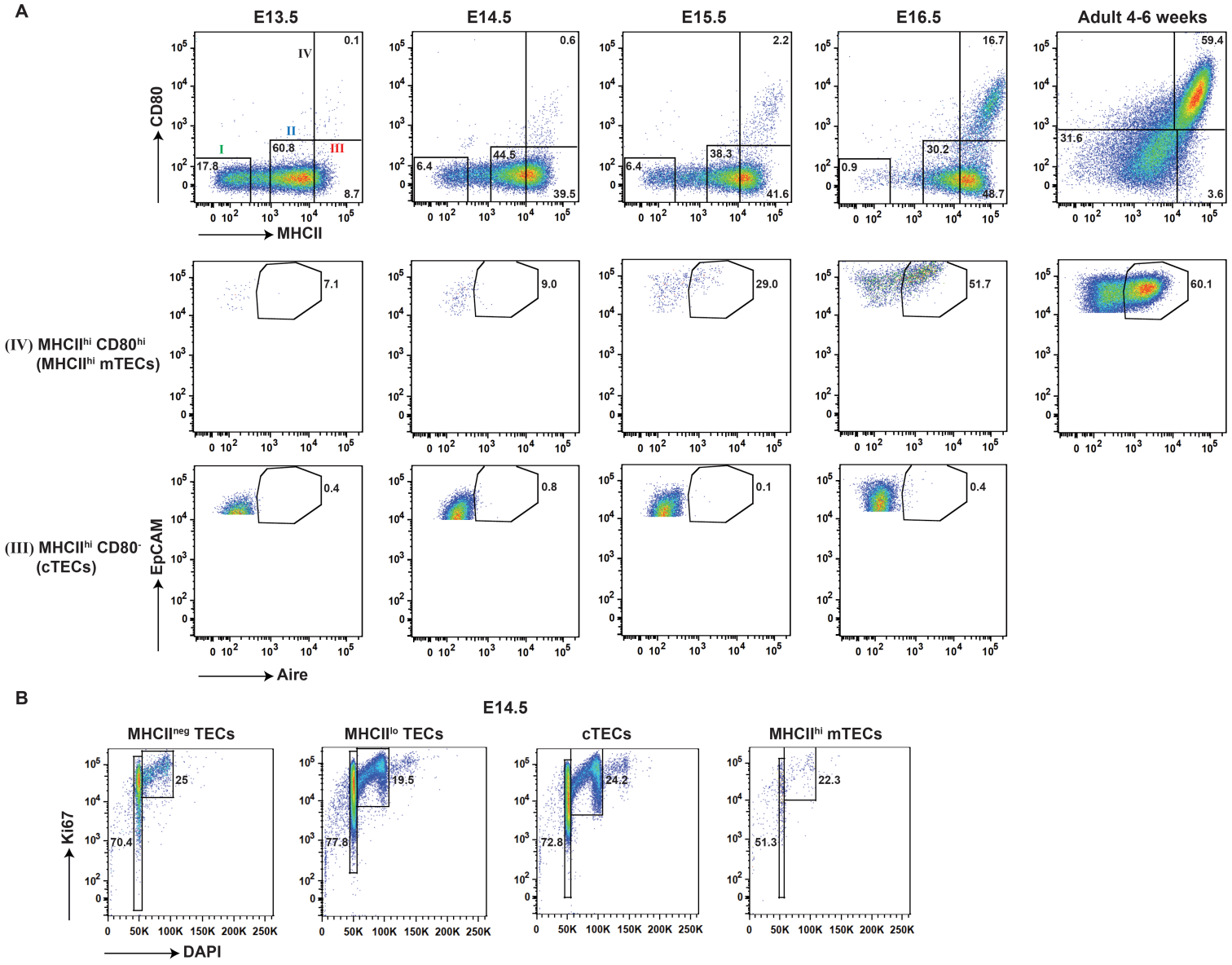
Dissecting and modeling the emergent murine TEC compartment during ontogeny

Supporting Information Figures



Supporting Information Figure 1

(A) The TEC identity of MHCII^{neg} was validated using the FoxN1-GFP reporter mouse. (B, C) Relative mRNA expression of mTEC (*Aire*, *RANK*, *Cld3* and *Ctss)* and cTEC lineage-specific genes (*β5t*, *Ctsl* and *Prss16)* in the four different TEC subsets at E15.5 (B) and p0 (C) were analysed. Results are representative of a pool of 7-12 C57BL/6 embryos/pups thymi. (D) Protein expression of the cTEC-specific surface marker Ly51 in the different TEC subsets at developmental stages E15.5 and p0.



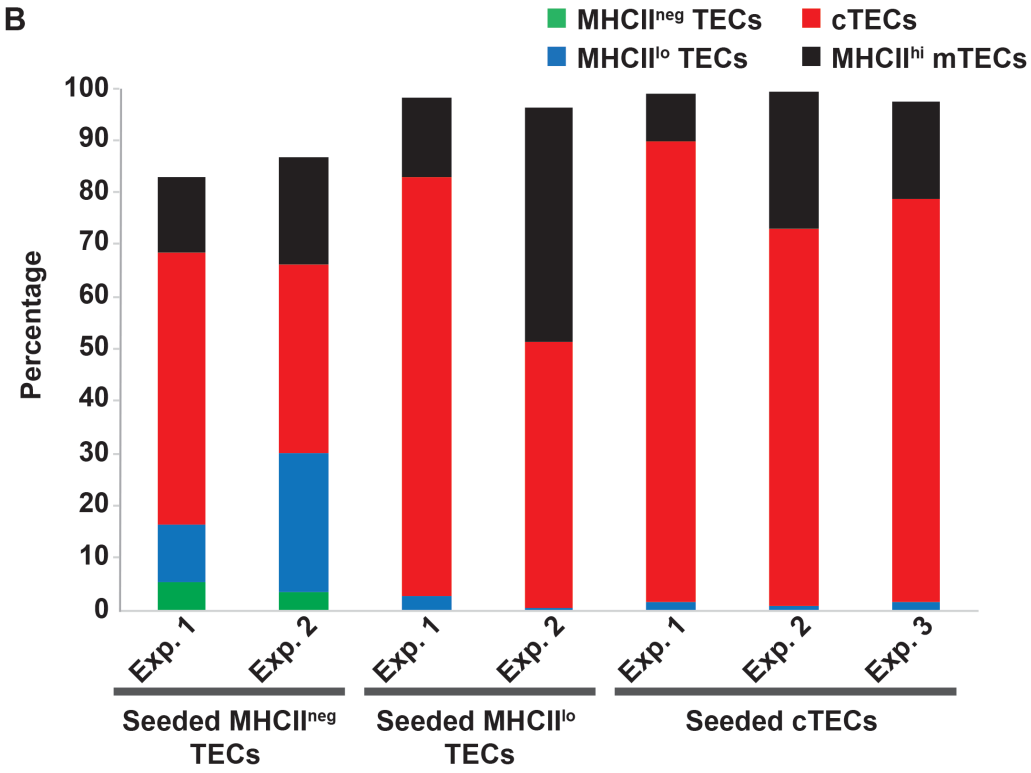
Supporting Information Figure 2

(A) Expression of Aire as a surrogate marker in MHCII^{hi}CD80^{hi}, but not in MHCII^{hi}CD80^{-}TECs during emrbyonic development. Results are representative of two to three independent experiments performed at the indicated time-points. Measurements for each experiment were based on 4-10 pooled embryonic or 2 adult thymi. (B) Representative dot plots of cell cycle analysis on E14.5 TEC subsets using Ki67 and DAPI. Cycling cells were defined as being Ki67$^{int/high}$ and DAPI^{+} 2N, comprising G2/M- and S- cell cycle phases.

A

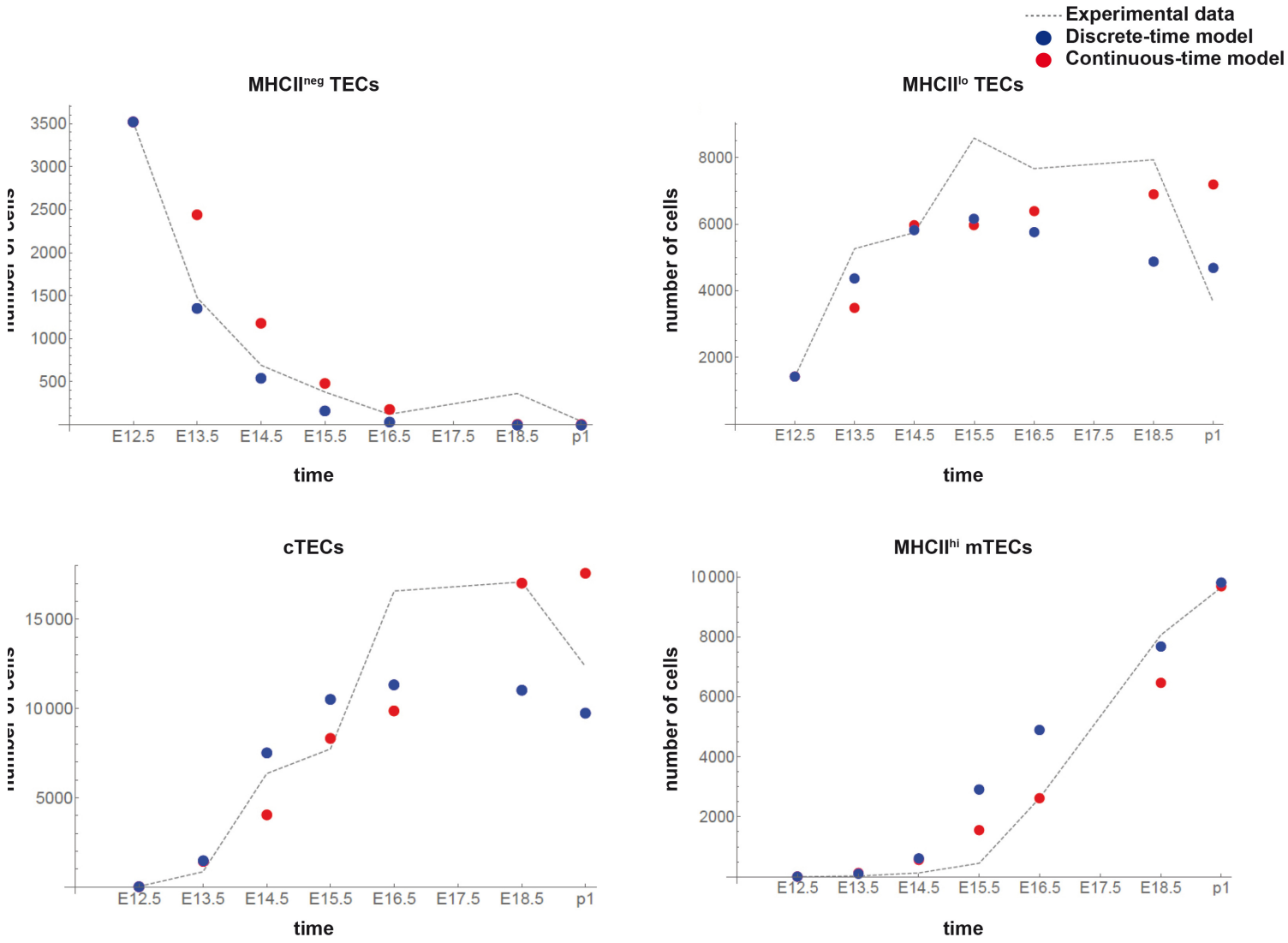
Seeded subset	Day 0			Day 4		
	Number of seeded cells			Percent CP450^{+} cells		
	Exp. 1	Exp. 2	Exp. 3	Exp. 1	Exp. 2	Exp. 3
MHCII^{neg} TECs	57000	41000	-	3.19	4.71	-
MHCII^{lo} TECs	47000	31000	-	6.23	4.17	-
cTECs	20000	39300	42000	20.3	23.5	32.2

B



Supporting Information Figure 3

(A) Sorted cell numbers for each TEC subsets seeded onto the RTOCs (day 0) and the percentage of CP450^{+} cells retrieved post-culture (day 4). (B) Percentage of each CP450^{+} TEC subset retrieved from two-three experiments on day 4.



Supporting Information Figure 4

Graphical comparison of the estimated number of all four TEC subsets as provided by the two proposed models (discrete- versus continuous-time models) compared to the average observed ex vivo TEC numbers.

Parameter estimates

Parameter	Discrete – time model		Continuous - time model	
	Estimate	95% CI	Estimate	95% CI
β	0.626	(0.569,0.655)	0.7	(0.667, 0.732)
δ	0.098	(0.084, 0.103)	0.095	(0.088 , 0.103)
λ_1	0.238	(0.209, 0.256)	0.233	(0.211, 0.254)
λ_2	0	0	0	0
λ_3	0.12	(0.092, 0.169)	0.135	(0.101, 0.168)
λ_4	0.06	(0.048, 0.56)	0.06	(0.043, 0.078)
λ_5	0	0	0	0
λ_6	0.03	(0.027, 0.034)	0.036	(0.033, 0.04)

Supporting Information Table 1

Parameter estimates and their confidence intervals for the discrete- versus continuous-time model setting.

FACS		
Antigen	**Antibody used (color, clone, company)**	**Dilution factor**
CD45	APC-eFluor780; 30-F11; eBioscience	1:200
Ly51	FITC; 6C3; BD Biosciences	1:100
EpCAM	PerCP-Cy5.5; G8.8; BioLegend	1:500
MHCII	Alexa647/Pacific Blue; M5/114.15.2; BioLegend/eBioscience	1:1000
CD80	PE; 16-10A1; BD Biosciences	1:100
Aire	FITC; B1/02-5H12-2; H. Scott, Adelaide, Australia	1:800
Ki67	PE Cy7; SolA15; ebioscience	1:200

Supporting Information Table 2

List of antibodies used in this study for flow cytometry analysis and sorting.

Supporting Information Methods

Mathematical models

The aim of this theoretical approach was to estimate a number of unknown parameters controlling the development of 4 TEC subsets during embryonic development. To describe the dynamics as shown in **Figure 3A**, we consider not only the number of cells of the given 4 cell types, but also the actually observed number of cells currently undergoing cell cycle for each cell type and embryonic day; those cell numbers are captured by the processes $N_1(t), N_2(t), N_3(t)$ and $N_4(t)$, and are known for each observed time point.

As a result, a total of 8 parameters have been estimated; for notation and interpretation of those parameters see **Figure 3A**. In particular, the parameters λ_1 to λ_6 describe the proportion of cells transforming into a different cell type per time unit. The parameter δ represents the proportion of cells dying through apoptosis per time unit. Finally, the parameter β represents the number of additional cells of the same subset produced by a cell currently undergoing cell cycle per time unit. Note that we assume both δ and β to be identical for all 4 TEC subsets. Additionally, we require all 8 parameters to be constant over time and to take values between 0 and 1, except for the parameter β, which is required to be non-negative only.

Discrete-time model:

In order to estimate the parameters, we first set up a discrete-time model. In the context of this model, given a known vector x_t of cell counts for the 4 cell types at a given time point t, the cell counts x_{t+1} at time point $t + 1$ can be predicted using a transformation matrix *M*, which takes the following form:

$$M = \begin{pmatrix} 1+\beta N_1(t)-\delta-\lambda_2-\lambda_1 & \lambda_2 & \lambda_1 & 0 \\ 0 & 1+\beta N_3(t)-\delta-\lambda_4-\lambda_6 & \lambda_4 & \lambda_6 \\ 0 & \lambda_3 & 1+\beta N_2(t)-\delta-\lambda_3-\lambda_5 & \lambda_5 \\ 0 & 0 & 0 & 1+\beta N_4(t)-\delta \end{pmatrix}$$

Each row and each column of this matrix corresponds to one of the 4 cell types: the first row/column represents MHCII^{neg} TECs; the second row/column represents cTECs; the third and fourth rows/columns represent MHCII^{lo} TECs and MHCII^{hi} mTECs, respectively. For *i* as the row index and *j* as the column index, the *(i, j)*- entry in the matrix *M* describes the effect the number of cells of type *i* at a given time point *t* has on the number of cells of type *j* at time point $t+1$. For example, the top left entry indicates that MHCII^{neg} TECs will increase in number through proliferation (as given by $+\beta\, N_1(t)$), but will also take losses through apoptosis (as given by $-\delta$) as well as transformation into cTECs (as given by $-\lambda_2$) and MHCII^{lo} TECs (as given by $-\lambda_1$). Similarly, the entry λ_2 in row 1, column 2 indicates that a proportion of $\lambda_2$of MHCII^{neg} TECs will transform into cTECs in one time unit.

Thus, for any given set of parameter estimate and starting cell number x, the resulting matrix *M* allows to make predictions for cell counts at future time points. In particular, the cell count x_{t+1} can then be estimated as $x_{t+1} = x_t * M$. Moreover, predictions over two or more time units can also be made: e.g., x_{t+2} can be derived as $x_{t+2} = x_{t+1} * M = x_t * M^2$. Note that all parameter estimates refer to the development in one time unit. We chose to set a time unit as 4 hours, i.e. 6 time units per embryonic day.

To fit the proposed discrete-time model to the experimental data, for each of the 7 available embryonic days we used a set of starting parameter values as well as the observed cell frequencies at this day as an input to create a prediction for the next day

(including a prediction for the first postnatal day p1 in case of E18.5). This prediction was then compared to the average of the actual cell numbers observed in the mice on the following embryonic day.

For each day and each cell type the squared deviation between observed and predicted values was computed and summed up over all cell subsets and days to obtain a goodness of fit criterion for any specific set of parameters. Then, parameter estimates minimizing this criterion were computed using the optimization procedure *"bobyqa"* from the *"minqa"* library in R Version 3.3.1.

Continuous-time model:

The main assumption of the discrete-time model above is the assumption that the cells differentiate, go into cell cycle or die at discrete times. To allow these processes to take place continuously, in the next step we also developed an alternative, continuous-time model for describing the same dynamics of the cell populations. This model is given by a system of differential equations where each equation represents the number of cells of each considered TEC subset at each moment of time.

In the context of the model presented in **Figure 3A**, we actually would have to set up a system with eight differential equations, i.e. one equation for each cell subset and an additional 4 equations for the processes $N_1(t), N_2(t), N_3(t)$ and $N_4(t)$. However, we did not have enough experimental data to fit such a system without an impact on the reliability of the obtained parameter estimates. To avoid possible inconsistency, we assumed that the number of cells in the division stage changes only at discrete times (in particular at the 7 available embryonic days) and thus the processes $N_1(t), N_2(t), N_3(t)$ and $N_4(t)$ are piecewise-constant within each embryonic day.

Under the latter assumption, we then obtained a system of only 4 differential equations, i.e. one equation for each cell subset. This system is as follows:

$$\begin{aligned} dX_1(t) &= \beta N_1(t) - (\delta + \lambda_1 + \lambda_2)\, X_1(t) \\ dX_2(t) &= \beta N_3(t) + \lambda_2\, X_1(t) + \lambda_3\, X_3(t) - (\delta + \lambda_4 + \lambda_6)\, X_2(t) \\ dX_3(t) &= \beta N_2(t) + \lambda_1\, X_1(t) + \lambda_4\, X_2(t) - (\delta + \lambda_3 + \lambda_5)\, X_3(t) \\ dX_4(t) &= \beta N_4(t) + \lambda_6\, X_2(t) + \lambda_5\, X_3(t) - \delta\, X_4(t) \end{aligned}$$

Wherein $N_1(t), N_2(t), N_3(t)$ and $N_4(t)$ is as in **Figure 3A**. Analogously to the discrete-time model above, each differential equation corresponds to one of the 4 cell subsets: the first differential equation describes the dynamics of the MHCII^{neg} TECs; the second the dynamics of the cTECs; the third the dynamics of the MHCII^{lo} TECs; and the fourth the dynamics of the MHCII^{hi} mTECs.

This model of differential equations was then fitted to the experimental data by means of the lowest square deviation criterion as introduced in case of the discrete-time model above. In particular, the functions *ParametericNDSolveValue* and *NonlinearModelFit* provided by *Mathematica 10* were applied while estimating the parameters. These functions allowed us to estimate the parameters numerically, without computing the analytical solution of the considered system of the differential equations.

We observed certain minor differences in the parameter estimates and their confidence intervals as presented in **Supporting Information Table 1**. Those differences are assumed to be a consequence of discrete- versus continuous- time model setting and the fact that the "*bobyqa"* optimization method, as applied in the discrete-time model, was not available for the continuous-time case.

Confidence intervals (CIs) for parameter estimates for both models were constructed by the non-parametric bootstrap method based on sampling with replacement for the

given experimental data set [1]. In each case, bootstrap samples were generated by stratified sampling, i.e. separate bootstrap samples were taken for each embryonic day and then combined into a full bootstrap data set; these CIs are presented in **Supporting Information Table 1** as well.

References

1. **Efron B**, **Tibshirani R**. *An introduction to the bootstrap*. Chapman & Hall/CRC Monographs on Statistics and Applied Probability; 1994.