Supplementary information for: Short-term Memory in Gene Induction Reveals the Regulatory Principle behind IL-4 Stochastic Expression

Author: Luca Mariani et al.

Table of Contents

Mathematical appendix 1

Post-transcriptional memory time

To understand the effect of mRNA degradation and protein secretion on the memory time of IL-4 expression, we consider the model presented in figure 2B of the main text; which includes only mRNA production-degradation, and protein translation-secretion:

$$\begin{aligned} \frac{dmRNA}{dt} &= k_R - d_R mRNA \\ \frac{dPROTEIN}{dt} &= k_P mRNA - d_P PROTEIN \end{aligned} \qquad \text{(Eq. S1.1)}$$

When the reactions happen stochastically, the model can be treated with two main approaches: the master equation, or the Langevin approach.

In the master equation approach, the system has a discrete number of states corresponding to the absolute values of the considered chemical compounds, and the system jumps between these states with a certain probability depending on the reaction rates [1]. When mRNA and proteins are considered separately, they follow a poissonian distribution, with the variance equal to the mean. In the mRNA case this leads to:

$$Var_{mRNA} \equiv < mRNA^2 > - < mRNA >^2 = < mRNA > = k_R / d_R \qquad \text{(Eq. S1.2)}$$

Conversely, in case that the mRNA is constant, the protein distribution verifies:

$$Var_{PROTEIN} = < PROTEIN > = k_R k_P / d_R d_P \qquad \text{(Eq. S1.3)}$$

When both of the compounds vary stochastically, the protein variance is perturbed by the upstream mRNA fluctuations [2], giving a variance:

$$Var_{PROTEIN} = < PROTEIN > (1 + \frac{k_P}{d_P + d_R}) \qquad \text{(Eq. S1.4)}$$

which leads to the Eq.1 in the main text, in which variance is normalized and $\tau_R \equiv d_R^{-1}, \tau_P \equiv d_P^{-1}$ are mRNA and protein averaged lifetimes.

In the Langevin approach, the system varies continuously in the space of the chemical compounds of mRNA and proteins. Then, random fluctuations are imposed to the kinetics of

the compounds by adding noise functions $\eta_{mRNA}, \eta_{PROTEIN}$ which perturb the equations of motion:

$$\begin{aligned} \frac{dmRNA}{dt} &= k_R - d_R mRNA + \eta_{mRNA}(t) \\ \frac{dPROTEIN}{dt} &= k_P mRNA - d_P PROTEIN + \eta_{PROTEIN}(t) \end{aligned} \qquad \text{(Eq. S1.5)}$$

In the case of white noises, the fluctuations are in average equal to zero and not correlated in time:

$$\begin{aligned} &< \eta_{mRNA}(t) >= 0 \ , \ < \eta_{mRNA}(t)\eta_{mRNA}(t+\tau) >= q_R \delta(\tau) \\ &< \eta_{PROTEIN}(t) >= 0 \ , \ < \eta_{PROTEIN}(t)\eta_{PROTEIN}(t+\tau) >= q_P \delta(\tau) \end{aligned} \qquad \text{(Eq. S1.6)}$$

where $\delta(\tau)$ is a Dirac-function, and the fluctuations depends just on the parameters q_R, q_P. The Langevin approach is consistent with the master equation when the gain-loss terms within the differential equations are linear, and the variance due to the noise functions η matches the variance obtained from the Master Equation. In the case of mRNA, the variance in the Langevin approach can be computed for the deviations from the steady-state, $\delta R = mRNA - < mRNA >$, by using the Fourier transform $x(t) = \int e^{i\omega t} x(\omega) d\omega / 2\pi$, as in [3]:

$$< \mid \eta_{mRNA}(\omega)^2 \mid >= q_R \ , \ \delta R(\omega) = \frac{\eta_{mRNA}(\omega)}{d_R + i\omega} \qquad \text{(Eq. S1.7)}$$

From the mRNA variance:

$$VAR_{mRNA} \equiv < \delta R^2 >= \int \frac{d\omega}{2\pi} \frac{q_R}{{d_R}^2 + \omega^2} = \frac{q_R}{2d_R} \qquad \text{(Eq. S1.8)}$$

and the "poissonian" constraint on the white noise intensity, $VAR_{mRNA} =< mRNA >= k_R / d_R$, it follows $q_R = 2k_R$.

To derive the value of q_P, a similar method is applied to the protein dynamics, while the mRNA is considered deterministic ($\eta_R = 0$), resulting in $q_P = 2k_P k_R / d_R$.

When the two noises are both present, the protein variance is perturbed as described by the master equation:

$$VAR_{PROTEIN} = \int \frac{d\omega}{2\pi} \frac{q_P}{{d_P}^2 + \omega^2} (1 + \frac{{k_P}^2 q_R}{{d_R}^2 + \omega^2}) = < PROTEIN > (1 + \frac{k_P}{d_P + d_R})$$

(Eq. S1.9)

equal to the variance in Eq. 1.4, obtained by Master Equation.

Following Sigal and collegues [4], the memory time of a certain component x can be defined as the half-time of the normalized autocorrelation function A_x:

$$A_x(t_{\frac{1}{2}}) = \frac{< \delta x(t) \delta x(t + t_{\frac{1}{2}}) >}{< \delta x(t)^2 >} = 0.5 \qquad \text{(Eq. S1.10)}$$

In the case of mRNA, the solution of δR reads:

$$\frac{d\delta R}{dt} = -d_R \delta R + \eta_{mRNA}(t); \qquad \delta R(t) = \int_{-\infty}^{t} e^{d_R(t'-t)} \eta_{mRNA}(t') dt' \qquad \text{(Eq. S1.11)}$$

Then, the normalized autocorrelation function of mRNA is:

$$A_R(\tau) = < \int_{-\infty}^{t} e^{d_R(t'-t)} \eta_{mRNA}(t') dt' \int_{-\infty}^{t+\tau} e^{d_R(t''-t-\tau)} \eta_{mRNA}(t'') dt'' > / VAR_{mRNA} =$$

$$= e^{-d_R(2t+\tau)} \int_{-\infty}^{t} e^{d_R t'} \int_{-\infty}^{t+\tau} e^{d_R t''} 2 d_R \delta(t' - t'') dt'' dt' = e^{-d_R \tau}$$

(Eq. S1.12)

Considering the estimated mRNA degradation rate for IL-4 (see Table 1), the memory time for mRNA is equal to $t_{\frac{1}{2}mRNA} \cong 1hr$.

Similarly, random variations in protein production, coupled to mRNA fluctuations, follow:

$$\frac{d\delta P}{dt} = k_P \delta R - d_P \delta P + \eta_{PROTEIN}(t)$$

$$\delta P(t) = \int_{-\infty}^{t} e^{d_P(t'-t)} \left\{ \eta_{PROTEIN}(t') + \int_{-\infty}^{t'} e^{d_R(t''-t')} \eta_{mRNA}(t'') dt'' \right\} dt' \qquad \text{(Eq. S1.13)}$$

with an autocorrelation function for protein levels equal to:

$$A_P(\tau)= \frac{d_R d_P k_P}{(d_P - d_R)(d_R + d_P + k_P)} \left(\frac{e^{-d_R \tau}}{d_R} - \frac{e^{-d_P \tau}}{d_P} \right) + \frac{d_R + d_P}{k_P + d_R + d_P} e^{-d_P \tau}$$

(Eq. S1.14)

In the case of IL-4, the estimate for the parameters ($d_P = 2d_R; k_P / d_R > 10^2$) leads to the following approximation:

$$A_P(\tau) \cong 2e^{-d_R \tau} - e^{-d_P \tau} \quad \text{(Eq. S1.15)}$$

Considering our estimate of the rates of protein secretion and mRNA degradation for IL-4 (Table 1), the theoretical memory time for IL-4 protein, if it would obey the model in Fig 1B, is equal to

$$t_{\frac{1}{2}PROTEIN} = 1.75hr \quad \text{(Eq. S1.16)}$$

Mathematical appendix 2

Transcription factor activity and occupancy of the binding site.

In the model in Fig 2C, the effect of the transcription factor activity on the gene regulation depends on the occupancy of the stimulation-dependent transcription factor (TF) at one specific binding site (BS).

The binding and the release of TF at the site is described by a simple mass action kinetics

$$TF + BS \underset{k^-}{\overset{k^+}{\rightleftarrows}} Complex(TF, BS)$$

where k^+ and k^- are the rates of binding and release respectively. The average time for a complex is equal to $B = 1/k^-$, while the average time in which the binding site is not bound by a TF is $N = 1/(TF\,k^+)$.

In the approximation of rapid equilibrium for TF binding-release, the occupancy of the site can be written as

$$O_{TF}(t) = \frac{B}{N + B} = \frac{TF(t)}{K_D + TF(t)} \quad \text{(Eq. S2.1)}$$

where $TF(t)$ is the kinetics of the TF concentration and $K_D = \frac{k^-}{k^+}$ is the dissociation constant.

For values of TF which do not saturate the occupancy of the binding site ($K_D >> TF(t)$), the occupancy in Eq. 2.1 is proportional to the transcription factor TF

$$O_{TF}(t) \cong \frac{TF(t)}{K_D} \qquad \text{(Eq. S2.2)}$$

Since $O_{TF}(t)$ is a multiplicative factor for the reactions of chromatin opening and mRNA transcription (Fig 2D), it can be arbitrarly re-scaled coordinatively with k_G and k_R . In the simulations presented in Fig. 2E, 3 and 6, the occupancy was rescaled relatively to the maximal value $O_{TF}(0) \equiv O_{TF}^{MAX} = 1$. In the simulations presented in Fig. 5, the value of O_{TF}^{MAX} was a free parameter in order to assess the effect of different values of TF activity on the gene expression.

Mathematical appendix 3

Fraction and mean expression value of IL-4 positive cells

To have analytical approximation on the fraction and the mean expression level of IL-4 positive cells on a population level, we considered just the initial constant phase of the kinetics for TF occupancy at the binding site $O_{TF}(t)$. This phase is equal to a step function with phase t_A:

$$\begin{aligned} &for\ \ 0 < t < t_A,\ \ O_{TF}(t) = O_{TF}{}^{Max} \\ &for\ \ t > \tau, \quad O_{TF}(t) = 0 \end{aligned} \qquad \text{(Eq. S3.1)}$$

Assuming that the binding site occupancy of the transcription factor switches off randomly and completely with probability rate δ_{TF}, the probability $S(t)$ of finding the cell with $O_{TF} = 0$ evolves as

$$\frac{dS}{dt} = \delta_{TF}(1 - S) \qquad \text{(Eq. S3.2)}$$

For a starting value $S(0) = 0$, the solution becomes

$$S(t) = 1 - e^{-\delta_{TF} t}. \qquad \text{(Eq. S3.3)}$$

and the probability distribution $D_{TF}(t)$ for a cell to have a phase *t* is equal to:

$$D_{TF}(t) = \frac{dS}{dt} = \delta_{TF} e^{-\delta_{TF} t} \ . \qquad \text{(Eq. S3.4)}$$

The mean phase of TF binding site occupancy, $< t_A >$, can be calculated as the expectation value of the each phase t_A over the probability distribution $D_{TF}(t)$ from Eq. S3.4:

$$< t_A > \equiv \int_0^\infty t_A D_{TF}(t_A) dt_A = \delta_{TF}^{-1} \qquad \text{(Eq. S3.5)}$$

Assuming that all genes are initally in the closed state, and they get opened in a single step with probability rate $\kappa_G O_{TF}(t)$, the probability $G(t)$ of a gene to open, evolves as:

$$\frac{dG}{dt} = k_G O_{TF}(t)(1-G) \qquad \text{(Eq. S3.6)}$$

and, depending on the phase τ, it is given by :

$$\begin{aligned} &for \ \ t < \tau, \quad G(t) = (1 - e^{-\kappa_G O_{TF}^{Max} t}) \\ &for \ \ t > \tau, \quad\ G(t) = (1 - e^{-\kappa_G O_{TF}^{Max} \tau}) \end{aligned} \qquad \text{(Eq. S3.7)}$$

Similarly to Eq. S3.4, for a cell with a phase τ, the probability distribution of the time elapsing from the beginning of the phase (t=0) to the opening of the gene $D_{ON}(t)$ is:

$$\begin{aligned} &for\ t < \tau, D_{ON}(t) = \frac{dG}{dt} = \kappa_G O_{TF}^{Max} e^{-\kappa_G O_{TF}^{Max} t} \\ &for\ t > \tau, D_{ON}(t) = \frac{dG}{dt} = 0 \end{aligned} \qquad \text{(Eq. S3.8)}$$

The fraction of opened genes at the end of the stimulation ρ_+ can be calculated as the probability to find a gene with a certain occupancy phase, distributed as $D_{TF}(\tau)$, activated within that phase $G(\tau)$, and using Eq. S3.4 and Eq. S3.7

$$\rho_+ \equiv \int_0^\infty D_{TF}(\tau) G(\tau) d\tau = \frac{< t_A > \kappa_G O_{TF}^{Max}}{1 + < t_A > \kappa_G O_{TF}^{Max}} . \qquad \text{(Eq. S3.9)}$$

Assuming that the the transcription rate for a closed gene is 0, while it depends on the binding site occupancy as $\kappa_R O_{TF}(t)$ for the open genes, the amount of mRNA molecules produced during a stimulation depends on the trascription phase, namely the time lapse between the gene opening and the end of the binding site occupancy. For a gene with occupancy phase τ, the mean transcription phase $\varphi(\tau)$ comes from the opening distribution $D_{ON}(t)$ in Eq. S3.8 as:

$$\varphi(\tau) \equiv \int_0^\infty D_{on}(t)(\tau - t)\, dt = \int_0^\tau \kappa_G O_{TF}^{Max} e^{-\kappa_G O_{TF}^{Max} t} (\tau - t) dt = \tau - \frac{1 - e^{-\kappa_G O_{TF}^{Max} \tau}}{\kappa_G O_{TF}^{Max}} \qquad \text{(Eq. S3.10)}$$

The mean transcription phase of the population, φ_{Total}, is equal to the expectation value of $\varphi(\tau)$ over the distribution of the occupancy phase $D_{TF}(\tau)$, and explicitely solved as:

$$\varphi_{Total} \equiv \int_0^\infty D_{TF}(\tau)\varphi(\tau)d\tau = < t_A > \frac{< t_A > \kappa_G O_{TF}^{Max}}{1 + < t_A > \kappa_G O_{TF}^{Max}}. \quad \text{(Eq. S3.11)}$$

To calculate the mean transcription phase of an opened gene φ_+, we rescale φ_{Total} with the fraction of opened genes at the end of the stimulation:

$$\varphi_+ \equiv \frac{\varphi_{Total}}{\rho_+} = < t_A >. \quad \text{(Eq. S3.12)}$$

The mean amount of protein produced by an opened gene, $\langle PROTEIN \rangle_+$, is indeed proportional to its mean trascrption time φ_+, to the transcriptional rate $O_{TF}^{Max} k_R$ and to the translational efficiency $\frac{k_P}{\delta_R}$:

$$\langle PROTEIN \rangle_+ = O_{TF}^{Max} \kappa_R < t_A > \frac{k_P}{\delta_R}. \quad \text{(Eq. S3.8)}$$

Supplementary figures legends

Supplementary Figure S1. Determination of the distribution for intracellular IL-4 positive cells.

(A) The mean fluorescence intensity (MFI) of the staining is the algebric mean of the fluorescence intensity distribution and σ is its standard deviation; in the depicted example MFI(IL-4)=26AU and σ(IL-4)=32AU. Typically, the distribution of the isotype antibody, which is used to detect the background auto-fluorescence, is shaped as the lower peak of the detection antibody distribution; in the depicted example MFI(ISO)=5AU and σ(ISO)=1.9AU . The normalized variance, $v = (\sigma/\mathrm{MFI})^2$, is computed and presented for the two distributions.
(B) To determine the exact percentage of IL-4 positive cells, the distribution of the isotype control (black curve) is re-scaled to overlap the lower peak of the normalized distribution of the detection IL-4 antibody (red curve). The overlapping area of the two distributions determines the fraction of cells which do not express IL-4, and the complementary area of the

IL-4 distribution quantifies the positive ones; in the depicted example 60% and 40% respectively.
(C) MFI, standard deviation and normalized variance of the positive cells are determined from the complementary area derived above (B).

Supplementary Figure S2. Accumulation rate of extracellular IL-4 peaks after ~4 hours from the beginning of the stimulation.

IL-4 accumulation kinetics in the supernatant was measured by ELISA upon re-stimulation with PMA/Ionomycin in wild-type Th2 cells. All measurements were normalized to the final accumulation value of each stimulation. For each single kinetics, the relative accumulation rate was determined as the increase in the normalized accumulation of extracellular IL-4 at each hour. Mean and standard deviation of 6 kinetics are presented.

Supplementary Figure S3. Estimation of the kinetics rates.

(A) To estimate experimentally τ_R, we used the kinetics of relative IL-4 mRNA upon stimulation that is also presented in Figure 1B (red circles). If mRNA degradation is not regulated but happens because of random encounters with specific enzymes ($-d_R * mRNA$), the mRNA lifetime distribution $D_{mRNA}(t) = e^{-d_R t}$ gives an average lifetime of mRNA equal to $\tau_R \equiv \int_0^\infty t\, D_{mRNA}(t) dt = d_R^{-1}$ and the mean value of mRNA molecules decays as $e^{-d_R t}$. Then, τ_R is the time needed by mRNA to decrease of a factor $e^{-1} = 0.37$ from its original value. An upper bound for the mRNA life time can be estimated from the decay time after the peak of the response, which is reached at hour 3. Since the mRNA relative signal takes ~120 minutes to decay from 1 to 0.37, 120 minutes would be the life time if mRNA production would stop abruptly after reaching the peak. Since such homogenous behavior in the population is unlikely, the true half-life most likely lies below this value. Analysis of later time points, when most cells have stopped transcription, yields the more realistic value between 90 and 60 minutes. These values are in agreement with previous reports [5], and we fixed an mRNA life time of 90 minutes for all the simulation and the model analysis.

(B) As in the mRNA case presented above, we assumed that protein secretion happens randomly ($-d_P * PROTEIN$), resulting in a protein average lifetime equal to $\tau_P \equiv \int_0^\infty t\, e^{-d_P t} dt = d_P^{-1}$. To estimate τ_P, we also used the data presented in Fig. 1B. We fitted the mRNA data with a spline (MatLab), which we used to fit the protein secretion rate to the kinetics of intracellular and secreted protein. In the fitting procedure, we used a model, where protein is produced in linear proportion to the mRNA level and secretion is described as a first-order process. The average protein kinetics is the solution of $\frac{d < PROTEIN >}{dt} = k_P < mRNA > -d_P * < PROTEIN >$. The best fit was found for a secretion rate of $d_P = 1.4\ hour^{-1}$, resulting in an average lifetime for intracellular protein of ~45 minutes. Profile-likelihood method [6] was used to find the 95% confidence interval (35min ,60min). In this method, for a parameter of interest h the log-increase $\phi(h)$ of the mean-square objective function $F(h)$, relative to the parameter minima $F(h_{\min})$, is calculated and rescaled to the number of fitting points (n) and the 95th quantile of the χ^2-distribution for one degree of freedom ($\chi^2_{1,0.95}$=3.841). As $\phi(h) = n\left|\ln(F(h) - \ln(F(h_{\min}))\right| / \chi^2_{1,0.95}$ reach the value 1, the 95th confidence interval is found.

Supplementary Figure S4. Data best-fitting for the values of the model parameters k_G, d_G and $<t_A>$.

(A) To find the best-fitting values of k_G, d_G and $<t_A>$ we compared the distributions of IL-4 positive cells from 2-weeks-differentiated wildtype Th2 cells (obtained as described in Fig. S1) with 5000 simulated cells with a conversion factor from protein number to fluorescence intensity of $10^{-3.7}$. We defined an objective function as the ordinary least-squares function of the positive fraction, mean fluorescence intensity and relative standard deviation at hour 2, 3, 4, and 6 between the considered distributions (12 values in total). With a simulated-annealing algorithm, we obtained values for k_G and t_A normally distributed (Lilliefors test). Profile-likelihood method, as described in Fig. S3B, confined the parameters in the region k_G =(0.16,0.4)h^{-1} and $<t_A>$=(1.1,1.5)h [6].

(B) Since d_G was not normally distributed and varied over several order of magnitude, we fixed it to an intermediate mean value, 0.02h^{-1}, and we varied systematically k_G and $<t_A>$ in the best-fit region. The minimum of the objective function were found for k_G=0.23h^{-1} and

$<t_A>$=1.3. The profile-likelihood based method was used to compute the 95% confidence intervals of the maximum likelihood parameter estimates, see Fig. S3B. The 95% confidence interval are k_G = (0.2,0.27)h^{-1} and $<t_A>$=(1.2,1.42)h.
(C) We applied the profile-likelihood method to estimate the upper bound for d_G=0.06h^{-1}.

Supplementary Figure S5. Short-term memory in IL-4 induction is cell-cycle independent.

(A) 1-week-differentiated *il4*^{wt}/*il4*^{gfp} Th2 cells were labeled with the proliferation cell marker DDAO and stimulated with anti-CD3/28 and costimulatory signals for 24 hours, and then rested, as described in Fig 4A. Either 3 or 5 days after first stimulation a fraction of cells were stimulated for 4 hours a second time with PMA/Ionomycin/BfA and stained for intracellular IL-4. As the cell divides, the two daughter cells exhibit a halved fluorescence intensity of the DDAO marker; therefore the level of cell proliferation is inversely correlated with the level of DDAO (see "Cell proliferation" arrow). Left panels: Intracellular IL-4 versus DDAO, measured by flow-cytometry, show that there is no correlation between IL-4 expression and cell proliferation in early (A) and late (B) second stimulations (Corr.Coef.<0.05). IL-4 expressing and non-expressing cells are outlined with red and blue dotted gates. Right panels: distributions of DDAO from IL-4 expressing (red line) and non-expressing (blue line) cells exhibit similar fluorescence profile. As high control for the DDAO labeling, resting cells at day 1 were measured (Black line, Day1). As low control, an aliquot of cells were left DDAO-unlabeled (grey shadow, no DDAO) and treated equally as the cells labelled with DDAO. One representative experiment out of two is shown.

Supplementary Figure S6. IL-4 Secretion assay.

1-week-differentiated *il4*^{wt}/*il4*^{gfp} Th2 cells were stimulated with anti-CD3/28 for 3.5h and the IL-4 secretion assay was performed. Right before the secretion phase, an aliquot of cells was removed and stored on ice (A, low control). To test for homogeneous matrix labeling, another

aliquot was incubated at a high cell density with recombinant IL-4 (30 ng/ml) at 37°C for 30 min (B, high control). The remaining cells were incubated at a low cell density for 30 min at 37°C (C). Surface-bound IL-4 was stained with a PE-conjugated antibody. Through cytometrix sorting the positive (D) and the negative (E) fractions were purified. To assure that surface labeling resulted from IL-4 expression in the same cell an aliquot of cells was fixed with formaldehyde before the sorting step (F+G). These cells were stained for intracellular IL-4 with an APC-coupled antibody (F), to test specificity of the staining, in parallel it was performed in the absence of saponin (G).

Supplementary Figure S7. The sorted IL-4 positive and negative populations do not express IL-4 after the stimulation period.

1-week-differentiated *il4*^{wt}/*il4*^{gfp} Th2 cells were stimulated, sorted for IL-4 and rested (see "Material and methods" in the main text). 24 hours after the termination of the stimulation, sorted IL-4 positive (red line) and negative (blue line) cells were stained for intracellular IL-4, and for isotype control (grey shadow).

Supplementary Figure S8. The IL-4 capture matrix on the cell surface interfered with the IL-4 intracellular staining for 2 days after the secretion assay.

Before the secretion assay (Fig. 4C and S6) some cells were removed and not labeled with the IL-4 capture matrix (unlabeled), but were otherwise treated identically to the rest of the cells (labeled). After the secretion phase and the surface IL-4 staining, cells were sorted, and then further stimulated and rested as described in Fig. 4C. The labeled, unlabeled, positive sorted, and negative sorted cells were stimulated 1, 2 or 3 days, as indicated, after the stimulation period. Stimulation was assessed for 4 hours with PMA/Ionomycin/BfA in the presence of the IL-4 specific antibody (24G2, 100µg/ml) that reduces the binding of IL-4 to the capture matrix present on the cell surface. After stimulation, the cells were fixed for intracellular staining and incubated with unlabeled IL-4 specific antibody (11B11) to saturate surface-bound IL-4. Then IL-4 was stained intracellularly with an APC-labeled IL-4 antibody (11B11,

BD) or with an isotype control antibody. At day 1 and 2 the staining signal was stronger in the labeled fraction compared to the unlabeled cells, although these samples should produce the same amount of IL-4, indicating that IL-4 on the cell surface bound to the capture matrix interfered with the intracellular staining (compare green and black lines). Therefore the capture matrix also resulted in a staining background in the sorted cells relative to the isotype controls (compare blue/red and grey) and intracellular staining could not be used at day 1 and 2 to compare IL-4 production in the positive and the negative fraction.

Supplementary Figure S9. Sorted IL-4 negative cells have similar probability of IL-4 induction than parental cells (Day0), and lower than the sorted IL-4 positive cells.

1-week-differentiated $il4^{wt}/il4^{gfp}$ Th2 cells were stimulated for 3.5 hours with CD3/CD28-specific antibodies, and IL-4 positive and negative cells were purified using the IL-4 secretion assay, followed by flow cytometric sorting. Sorted IL-4 positive and negative cells were further stimulated for an additional 20 hours and rested (see "Material and methods" of the main text). After 3, 4 or 5 days in resting conditions, the cells were re-stimulated with PMA/Ionomycin/BfA, and IL-4 production was measured by intracellular staining followed by flow cytometry. The percentage of IL-4 producing cells in each population was quantified as described in Fig. S1. Left panels: IL-4 expression profiles for five independent repetitions (exp.1-5) are presented and used for statistical analysis of the results. Right panel: For each single repetition, the percentage of IL-4 positive cells at day 3, day 4, and day 5, relative to the value at day 0, are presented and used to compute the statistical mean and standard deviation presented in Fig 4C. By applying a t-test, significant differences were found among the sorted IL-4 positive cells at different days (Matlab, P-value<0.05), and between the sorted IL-4 positive and negative at different days. On contrary, the sorted IL-4 negative cells were not significantly different (Matlab, P-value<0.05) from a normal distribution with mean value 1, which is the reference value of day 0.

Supplementary Figure S10. In the first days after the initial stimulation, IL-4 extracellular accumulation was higher in the sorted IL-4 positive population than in the negative population.

1-week-differentiated *il4*^{wt}/*il4*^{gfp} Th2 cells were stimulated for 3.5 hours with CD3/CD28-specific antibodies, and IL-4 positive and negative cells were purified using the IL-4 secretion assay, followed by flow cytometric sorting (see fig.S6). Sorted IL-4 positive and negative cells were stimulated for an additional 20 hours and rested, (see "Material and methods" of the main text). After 1 or 2 days in resting conditions, fractions of cells were re-stimulated with PMA/Ionomycin for 20 hours and IL-4 production was measured in the supernatant by ELISA. The sorted IL-4 positive cells produced significantly more IL-4 than the negative cells. Moreover, IL-4 production in the positive cells declined from day 1 to day 2 (loss of memory) whereas it remained similar in the negative fraction. One representative experiment out of two is shown.

Supplementary Figure S11. In *il4*^{wt}/*il4*^{gfp} Th2 cells, the expression of *il4*^{gfp} allele is similar in cell populations previously sorted according to expression or non-expression of the *il4*^{wt} allele.

1-week-differentiated *il4*^{wt}/*il4*^{gfp} Th2 cells were stimulated with anti-CD3/28 for 3.5 hours. Cells producing and not producing IL-4 from the *il4*^{wt} allele were purified using IL-4 secretion assay (Material and Methods). Sorted IL-4 positive and negative cells were stimulated for an additional 20 hours and rested, as described in Fig. 4C and in Material and Methods. At day 1, day 2, and day 3 after stimulation, aliquots of sorted IL-4 positive (red line) and negative (blue line) cells were stimulated with PMA/ionomycin for 20 hours and measured with flow-cytometry. To check that GFP was expressed *de novo*, aliquots of cells were left unstimulated and measured, verifying that GFP was not continuously expressed (day

1 before stimulus), and any expression seen on subsequent days was due to the stimulation of the cells. GFP expression was similar in the sorted IL-4 producers and non-producers, sometimes with slightly higher GFP expression in the sorted IL-4 producers than in the non-producers (e.g., day 2). By comparison, expression of the $il4^{wt}$ allele was strongly enhanced in the sorted IL-4 producers, as shown for aliquots of cells stimulated on day 3 for 4 hours with PMA/ionomycin/BfA and stained for intracellular IL-4 (see also Figure S9). An independent repetition of this experiment also yielded similar or slightly elevated GFP expression in sorted IL-4 producers versus non-producers, as compared to the strong enrichment of the sorted IL-4 producers in expression of the $il4^{wt}$ allele.

Supplementary Figure S12. Sorted IL-4 positive and negative cells do not differ in proliferation rate.

1-week-differentiated $il4^{wt}/il4^{gfp}$ Th2 cells were stimulated with anti-CD3/28 for 3.5 hours. IL-4 producing and non-producing cells were purified, using IL-4 secretion assay (see ”Materials and Methods” in main text), labelled with DDAO. Sorted IL-4 positive and negative cells were stimulated for additional 20 hours and rested (see “Material and methods” of the main text). After 3 and 5 days of resting culture, DDAO fluorescence intensity were measured with flow-cytometry in both populations. As the cell divides, the two daughter cells have an halfed fluorescence intensity of the DDAO marker, therefore the level of cell proliferation is inversely correlated with the level of DDAO (see “cell proliferation” arrow). Distribution of the proliferation marker DDAO in sorted IL-4 positive (red line) and negative (blue line) cells exhibit similar fluorescence profiles. As control, an unlabelled population treated with same conditions is presented (grey shadow). One representative experiment out of two is shown.

Supplementary Figure S13. Sorted IL-4 positive and negative populations exhibit similar up-regulation of Gata-3 expression.

1-week-differentiated *il4*^{wt}/*il4*^{gfp} Th2 cells were stimulated with anti-CD3/28 for 3.5 hours. IL-4 producing (A) and non-producing (B) cells were purified with IL-4 secretion (see fig. S6) and stimulated for an additional 20 hours and rested (see "Material and methods" of the main text). After 5 days in resting conditions, Gata-3 was measured in sorted IL-4 positive (red line) and negative cultures (blue line). One representative experiment out of four is shown (two repetitions using heterozygous *il4*^{wt}/*il4*^{gfp} mice [7] and two repetitions wild-type mice).

Supplementary Figure S14. Confidence intervals for the chromatin opening rate in 1-week- and 3-weeks-differentiated wild-type Th2 cells.

We used the same objective function F, as described in Fig. S4, and the same parameters (Table 1) to fit data from (A) 1-week- and (B) 3-weeks-differentiated wild-type Th2 cells with the model simulation, where the chromatin opening rate k_G was the unique free parameter. Profile-likelihood based method was used to compute the 95% confidence intervals of the chromatin opening rate k_G for the two sets of data.

Supplementary Figure S15. Correlation between positive fraction and mean protein level in case of monoallelic regulation.

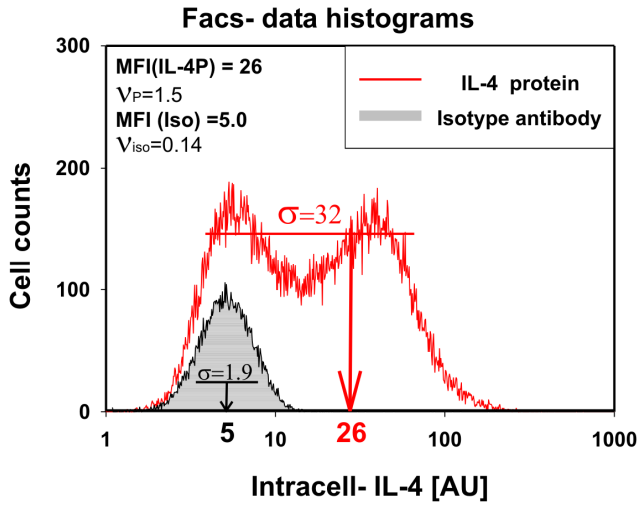
The model considers a population of cells carrying two alleles of a certain gene. Upon stimulation, each allele activates its expression independently of the other one with probability rate ρ_1 and a mean protein level P_1. The fraction of cells expressing one allele (dark grey area) is $\rho_M = 2\rho_1(1-\rho_1)$ with mean level P_1. The fraction expressing two alleles (light grey area) is $\rho_B = \rho_1^{\,2}$ with mean level $2P_1$. The total ρ_T fraction of expressing cells (horizontal axis) is the sum of $\rho_M + \rho_B = \rho_T = \rho_1(2-\rho_1)$ and the mean protein level is $P_T = P_1\rho_M + 2P_1\rho_B = 2\rho_1 P_1$. The mean protein level per expressing cell is $P_+ = P_T/\rho_T = 2P_1(1-\sqrt{1-\rho_T})/\rho_T$. For a constant production P_1 from each allele, an induction of $\rho_T = 12\%$ of the population (arrow on 0.12) implies a mean protein level in the expessing fraction of $P_+(12\%) = 1.03P_1$, while an induction of $\rho_T = 60\%$ (arrow on 0.6) implies a mean

protein level in the expessing fraction of $P_+(60\%) = 1.22P_1$. Compared to the basic value of a single allele P_1, the increases in the protein level $(P_+ - P_1)/P_1$ (black dotted line) for the two fractions were of 3% and 22% respectively. Between these two production levels, the relative increase of the mean level was ~18%, as presented in the main text.

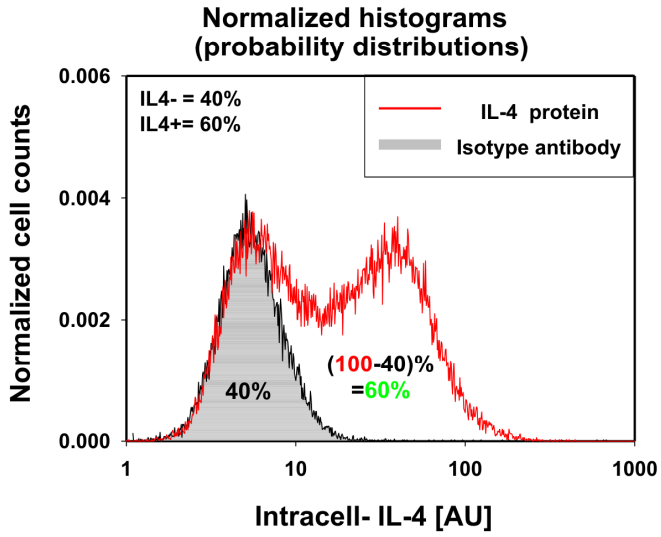
Supplementary information references

1. Van Kampen NG (2004) Stochastic processes in physics and chemistry. Amsterdam: Elsevier. 465 S. p.
2. Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A 98: 8614-8619.
3. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. Nat Genet 31: 69-73.
4. Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, et al. (2006) Variability and memory of protein levels in human cells. Nature 444: 643-646.
5. Yarovinsky TO, Butler NS, Monick MM, Hunninghake GW (2006) Early exposure to IL-4 stabilizes IL-4 mRNA in CD4+ T cells via RNA-binding protein HuR. J Immunol 177: 4426-4435.
6. Verzon D, Moolgavkar SH (1988) An algorithm for computing profile-likelihood-based confidence intervals. Appl Stat 37: 87-94.
7. Hu-Li J, Pannetier C, Guo L, Lohning M, Gu H, et al. (2001) Regulation of expression of IL-4 alleles: analysis using a chimeric GFP/IL-4 gene. Immunity 14: 1-11.

A



B



C

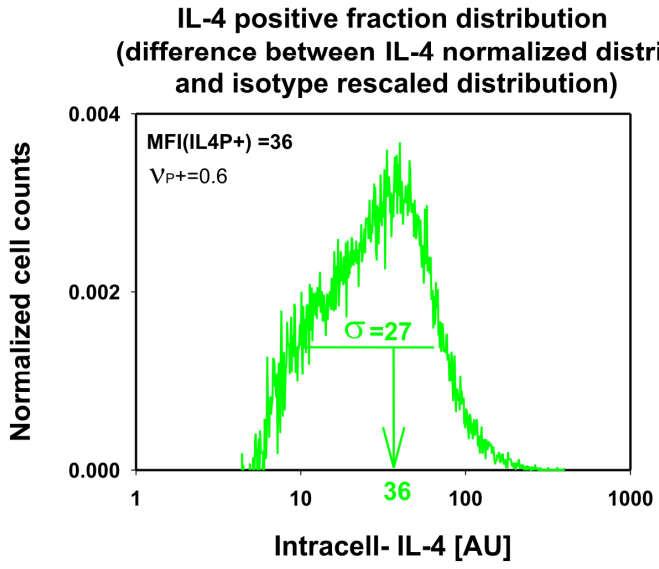


Figure S1

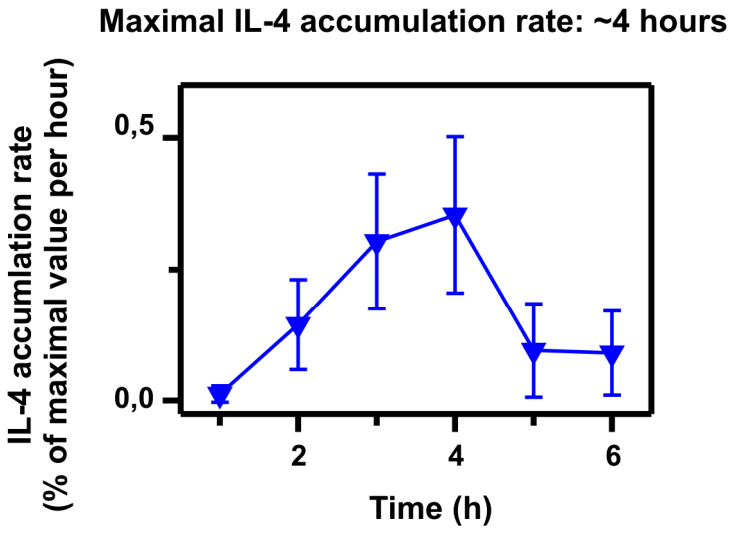
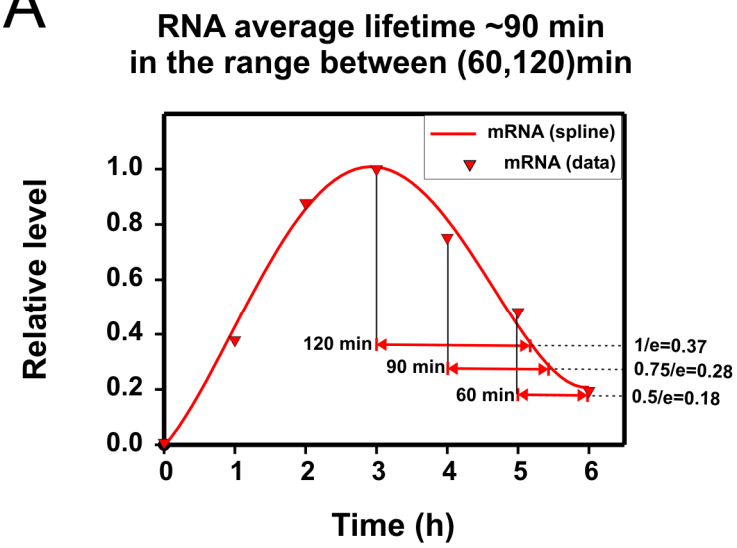

Maximal IL-4 accumulation rate: ~4 hours
IL-4 accumulation rate
(% of maximal value per hour)
0,5
0,0
2
4
6
Time (h)

Figure S2

A


RNA average lifetime ~90 min
in the range between (60,120)min
mRNA (spline)
mRNA (data)
Relative level
1.0
0.8
0.6
0.4
0.2
0.0
120 min
90 min
60 min
1/e=0.37
0.75/e=0.28
0.5/e=0.18
0
1
2
3
4
5
6
Time (h)

B

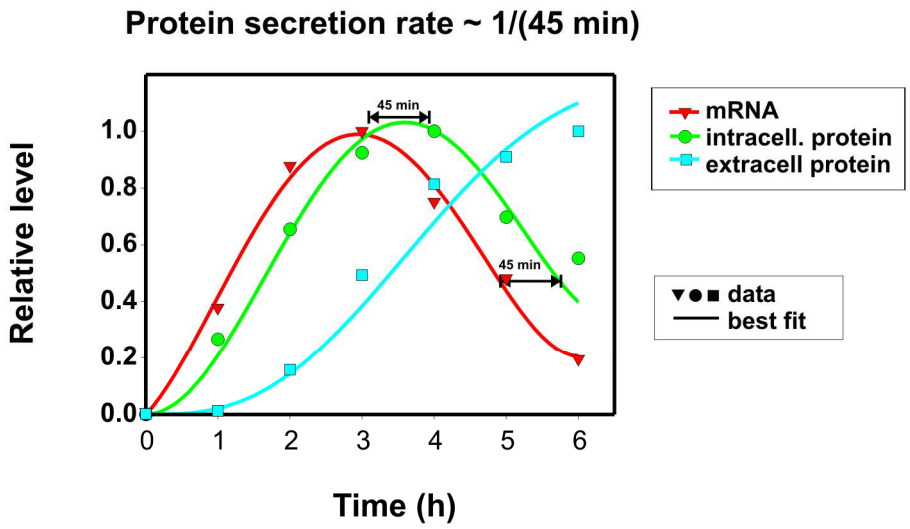
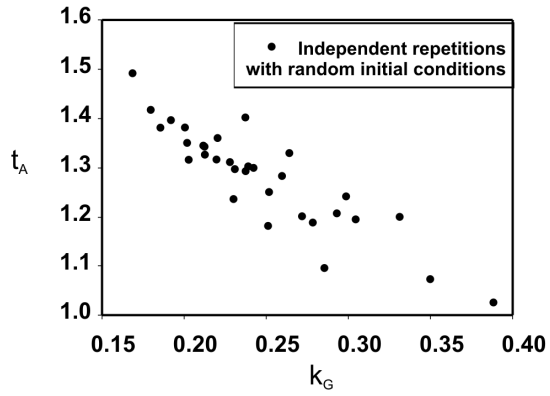

Protein secretion rate ~ 1/(45 min)
Relative level
1.0
0.8
0.6
0.4
0.2
0.0
45 min
45 min
mRNA
intracell. protein
extracell protein
data
best fit
0
1
2
3
4
5
6
Time (h)

Figure S3

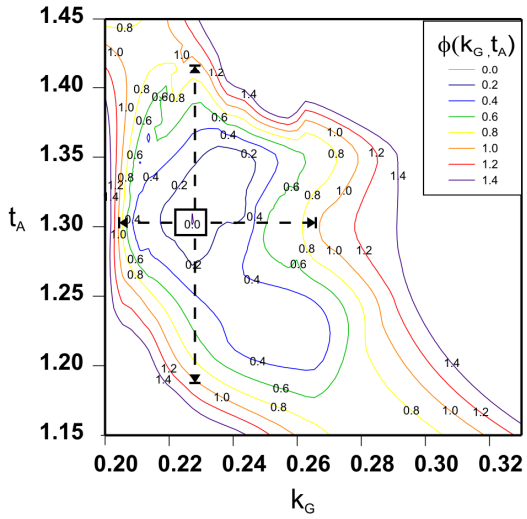
A

Best fit regions of k_G and t_A
determined from simulated annealing



B

Profile-likelihood method for k_G and t_A
around (k_G=023/h,t_A=1.3h,d_G = 0.01/h)



C

Profile-likelihood method for d_G
around (k_G=023/h,t_A=1.3h,d_G = 0.01/h)

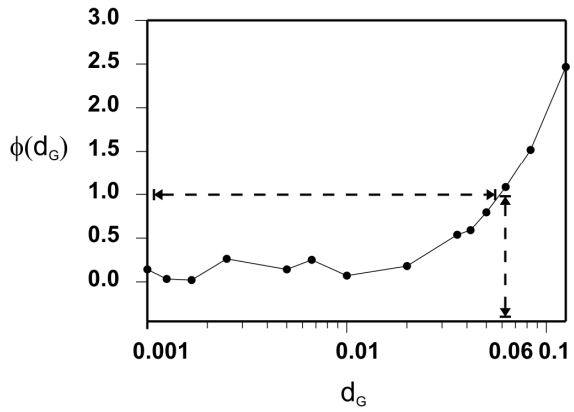
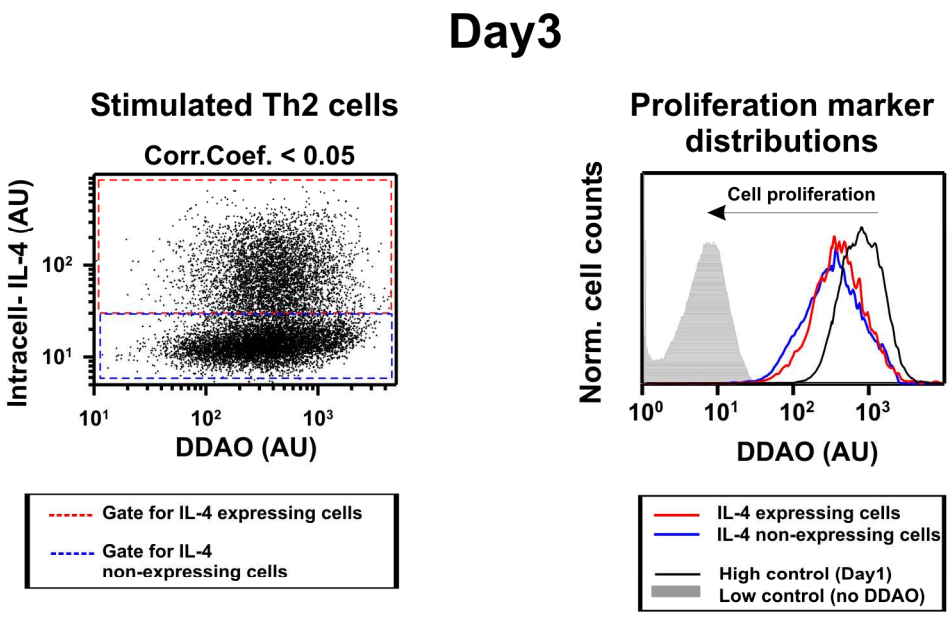


Figure S4

A


Day3
Stimulated Th2 cells
Corr.Coef. < 0.05
Intracell- IL-4 (AU)
DDAO (AU)
Gate for IL-4 expressing cells
Gate for IL-4 non-expressing cells
Proliferation marker distributions
Norm. cell counts
Cell proliferation
DDAO (AU)
IL-4 expressing cells
IL-4 non-expressing cells
High control (Day1)
Low control (no DDAO)

B

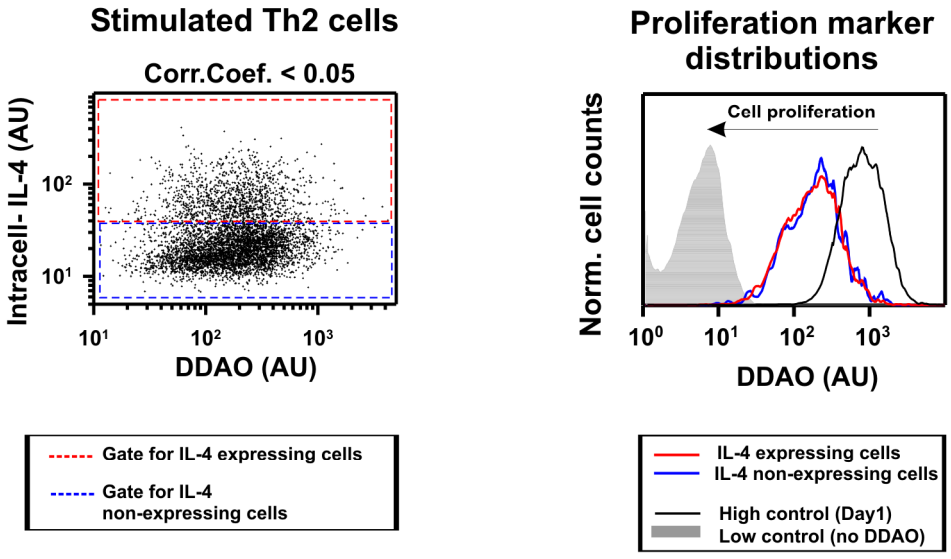

Day5
Stimulated Th2 cells
Corr.Coef. < 0.05
Intracell- IL-4 (AU)
DDAO (AU)
Gate for IL-4 expressing cells
Gate for IL-4 non-expressing cells
Proliferation marker distributions
Norm. cell counts
Cell proliferation
DDAO (AU)
IL-4 expressing cells
IL-4 non-expressing cells
High control (Day1)
Low control (no DDAO)

Figure S5

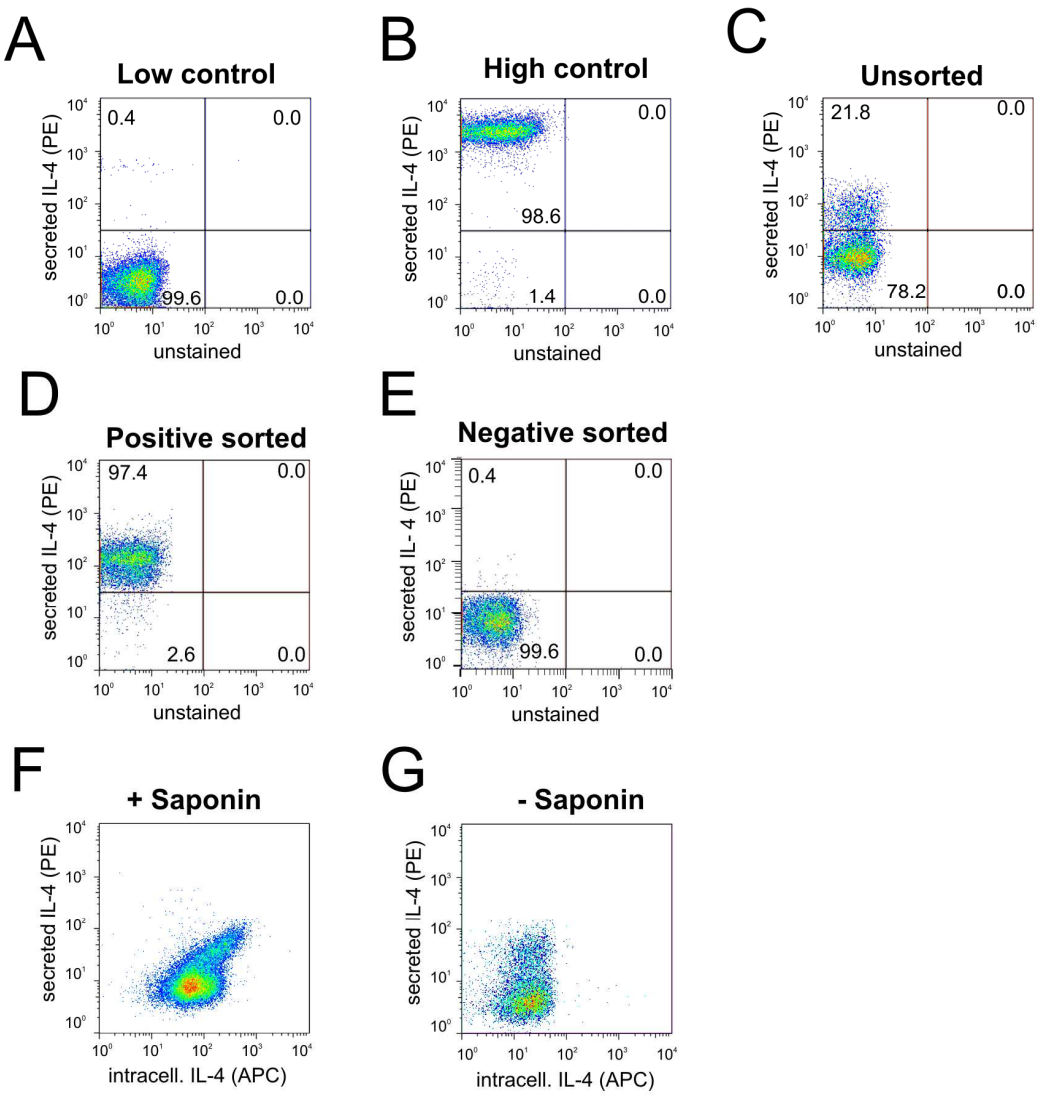

A
Low control
secreted IL-4 (PE)
unstained
0.4
0.0
99.6
0.0
B
High control
secreted IL-4 (PE)
unstained
0.0
98.6
1.4
0.0
C
Unsorted
secreted IL-4 (PE)
unstained
21.8
0.0
78.2
0.0
D
Positive sorted
secreted IL-4 (PE)
unstained
97.4
0.0
2.6
0.0
E
Negative sorted
secreted IL- 4 (PE)
unstained
0.4
0.0
99.6
0.0
F
+ Saponin
secreted IL-4 (PE)
intracell. IL-4 (APC)
G
- Saponin
secreted IL-4 (PE)
intracell. IL-4 (APC)

Figure S6

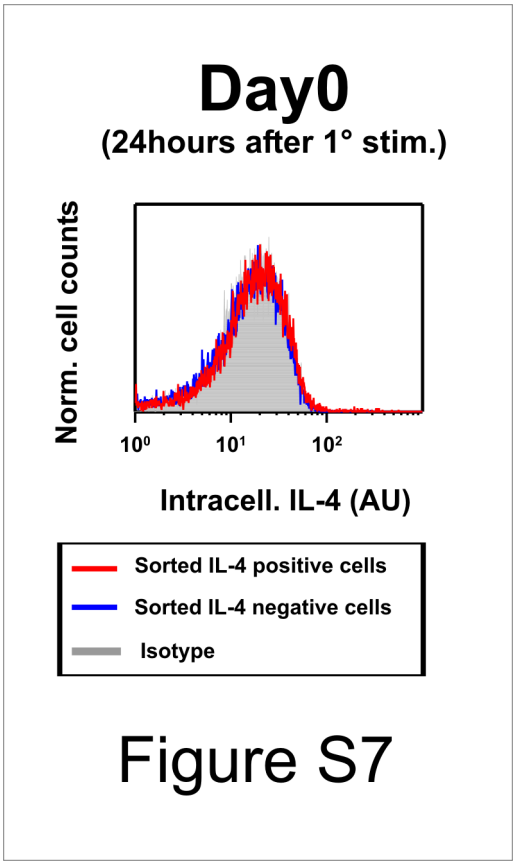

Day0
(24hours after 1° stim.)
Norm. cell counts
10⁰
10¹
10²
Intracell. IL-4 (AU)
Sorted IL-4 positive cells
Sorted IL-4 negative cells
Isotype

Figure S7

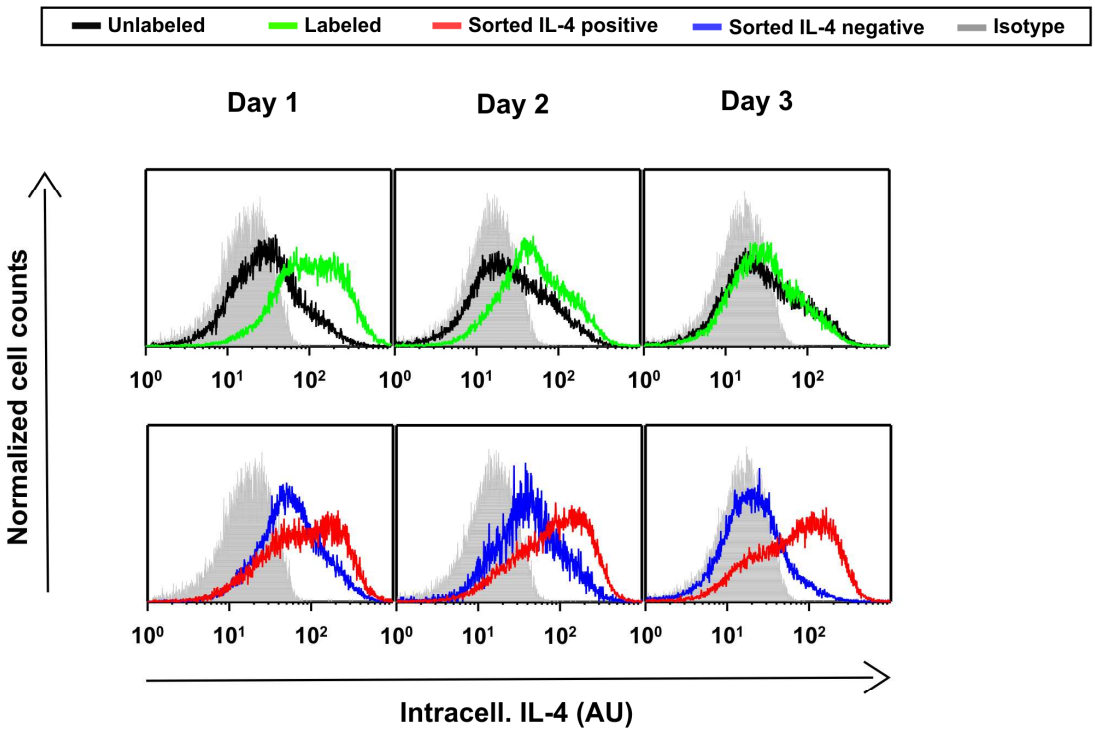

Unlabeled
Labeled
Sorted IL-4 positive
Sorted IL-4 negative
Isotype
Day 1
Day 2
Day 3
Normalized cell counts
10⁰ 10¹ 10² 10⁰ 10¹ 10² 10⁰ 10¹ 10²
10⁰ 10¹ 10² 10⁰ 10¹ 10² 10⁰ 10¹ 10²
Intracell. IL-4 (AU)

Figure S8

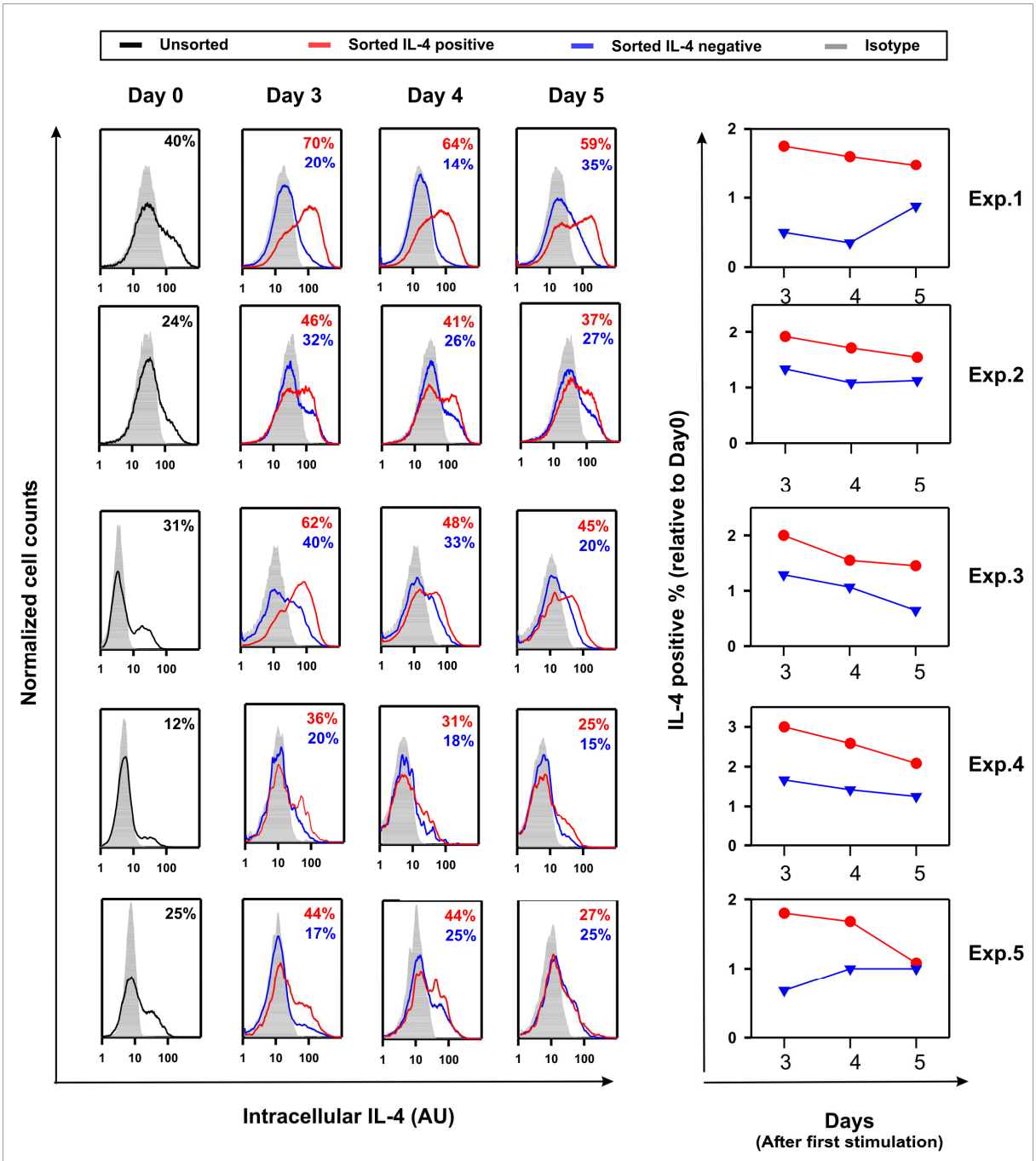

Unsorted
Sorted IL-4 positive
Sorted IL-4 negative
Isotype
Day 0
Day 3
Day 4
Day 5
40%
70%
20%
64%
14%
59%
35%
24%
46%
32%
41%
26%
37%
27%
31%
62%
40%
48%
33%
45%
20%
12%
36%
20%
31%
18%
25%
15%
25%
44%
17%
44%
25%
27%
25%
Normalized cell counts
Intracellular IL-4 (AU)
IL-4 positive % (relative to Day0)
Days
(After first stimulation)
Exp.1
Exp.2
Exp.3
Exp.4
Exp.5

Figure S9

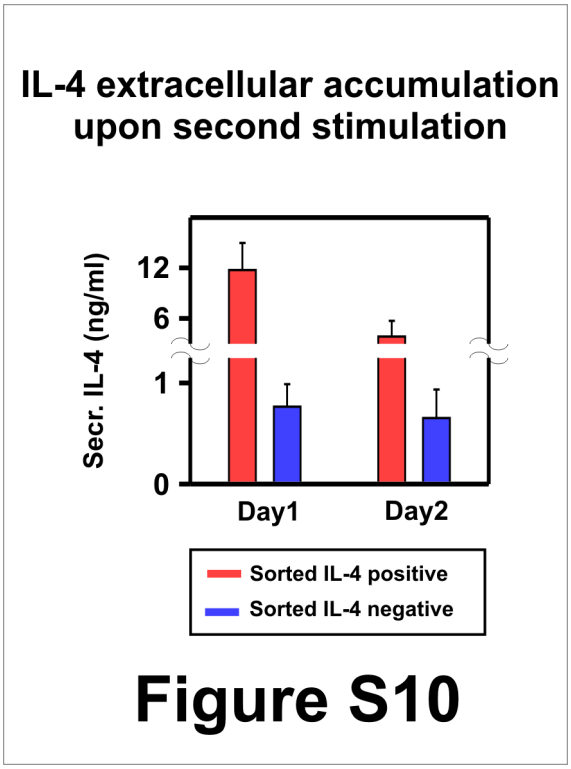


Figure S10

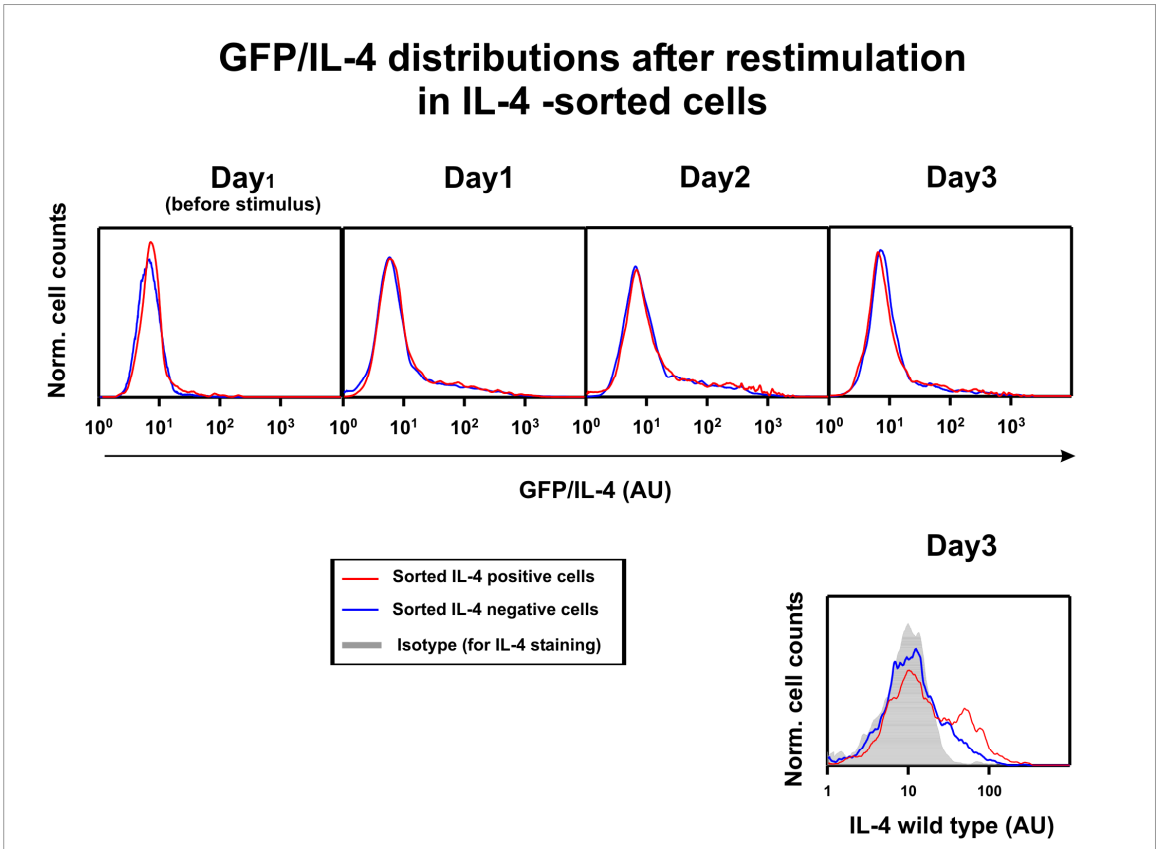


Figure S11

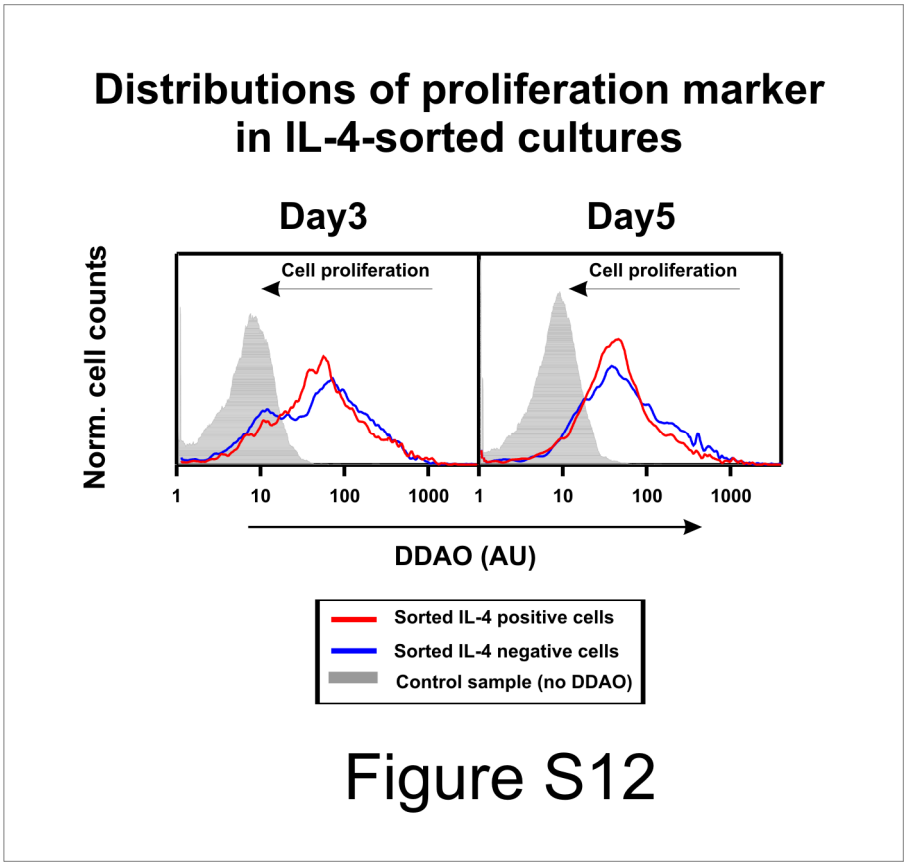


Figure S12

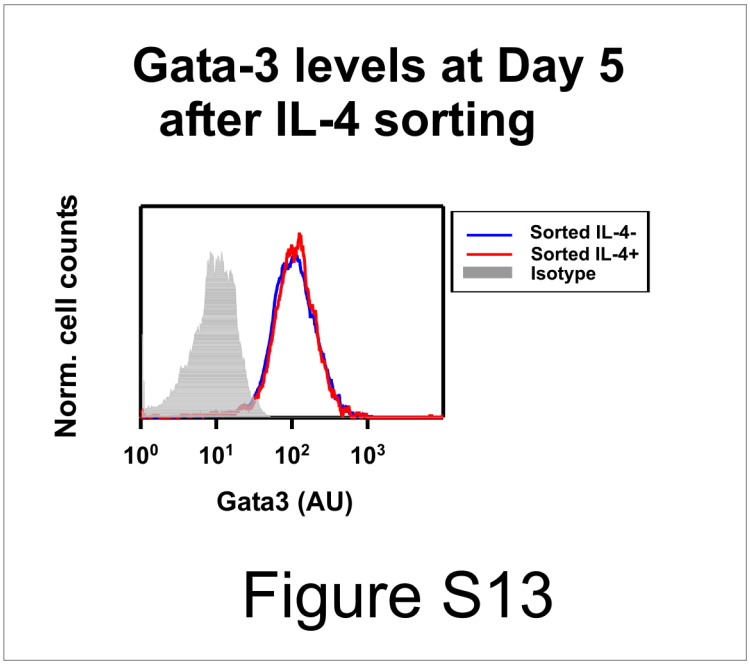


Figure S13

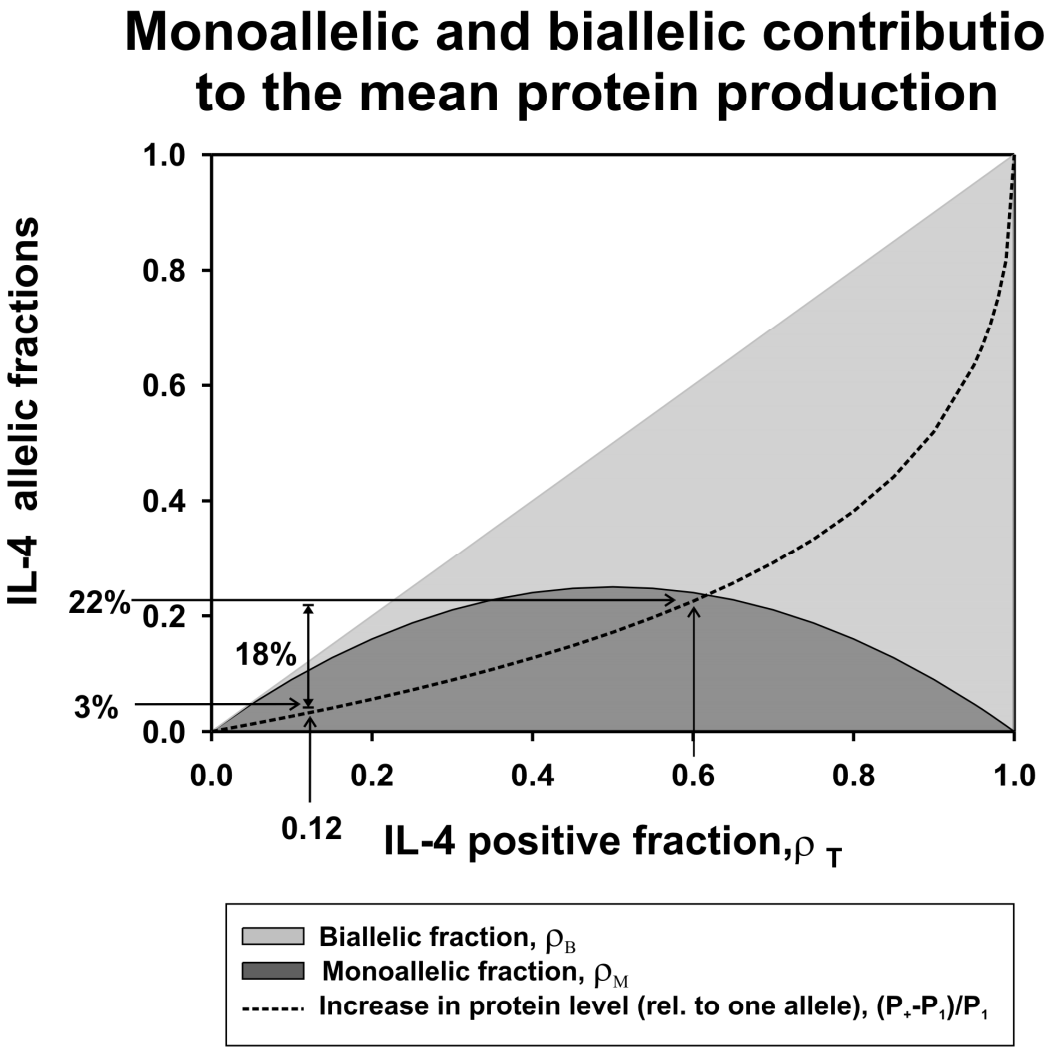
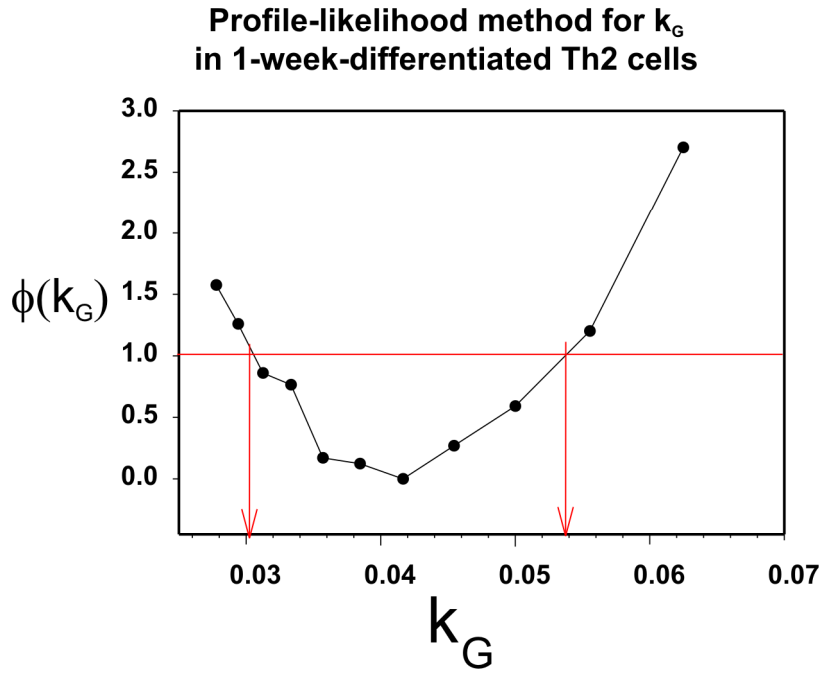

Monoallelic and biallelic contribution to the mean protein production
IL-4 allelic fractions
1.0
0.8
0.6
0.4
0.2
0.0
22%
18%
3%
0.0
0.2
0.4
0.6
0.8
1.0
0.12
IL-4 positive fraction, ρ T
Biallelic fraction, ρB
Monoallelic fraction, ρM
Increase in protein level (rel. to one allele), (P+-P1)/P1

Figure S14

A



B

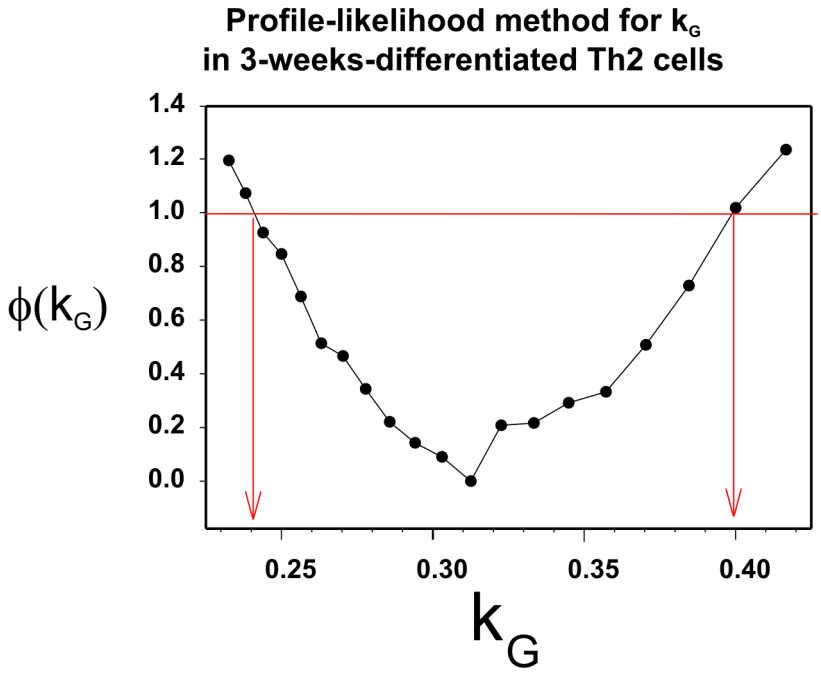


Figure S15