Supplementary Material for "Differentiation of developing olfactory neurons analyzed in terms of coupled epigenetic landscapes"

Anne Katrine Alsing* and Kim Sneppen**

Center for Models of Life, Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen Ø

* *alsing@nbi.dk*

** *corresponding author; sneppen@nbi.dk*

S1 - Stability analysis of the model

To analyze eq. 4 in the main text we consider two single genes, with active nucleosome fractions a and b, to be freely varying in a situation where all other $N-2$ genes are slaved to a single variable active nucleosome fraction, c.

$$\begin{aligned}\frac{da}{dt} &= \frac{1}{1+r(a^h+b^h+(N-2)c^h)}\cdot a^2\cdot(1-a)-\mu\cdot a\cdot(1-a)^2 \\ &- \beta\cdot a+\frac{\beta(1-a)}{1+r(a^h+b^h+(N-2)c^h)} \qquad (1)\\ \frac{db}{dt} &= \frac{1}{1+r(a^h+b^h+(N-2)c^h)}\cdot b^2\cdot(1-b)-\mu\cdot b\cdot(1-b)^2 \\ &- \beta\cdot b+\frac{\beta(1-b)}{1+r(a^h+b^h+(N-2)c^h)} \qquad (2)\\ \frac{dc}{dt} &= \frac{1}{1+r(a^h+b^h+(N-2)c^h)}\cdot c^2\cdot(1-c)-\mu\cdot c\cdot(1-c)^2 \\ &- \beta\cdot c+\frac{\beta(1-c)}{1+r(a^h+b^h+(N-2)c^h)} \qquad (3)\end{aligned}$$

Main text Fig. 3 B,E,H, show the sign of da/dt and db/dt assuming that c takes the lowest value where $dc/dt = 0$. This assumption is reasonable

because $dc/dt > 0$ for $c = 0$ and for increasing c the first null cline should therefore be stable against variations in c.

From eq. 3 we can deduce the fraction of active nucleosomes in repressed genes, as this is reflected in the lowest c value where $dc/dt = 0$. This "nucleosomal noise" is approximately $c = \beta/(\mu(1+r)) \sim \beta$. Using main text eq. 1, this minority fraction of active nucleosomes is associated to a minority production from each repressed gene of about β^h. For comparison, the receptor production from the chosen active gene is $a^h \sim 1$ where a is determined from largest a that fulfils eq. 1. This a value correspond to the "$a \sim 1$" solution in the lowest right corners of main text Fig. 3B,E & H. Requiring that the cell response is dominated by the chosen gene, imply that the activity of the chosen gene should be in excess of all the remaining N-1 genes together, $a \sim 1 > N \cdot \beta^h$, a demand that is fulfilled for all examined parameters.

S2 - Robustness to model parameters

The investigation presented in the main text uses standard parameters, with $N = 100$ genes each covered by $L = 50$ nucleosomes that interact with parameters $\beta = 0.03$, $r = 1$ and hill coefficient $h = 2$. Here we demonstrate that varying these parameters leave our main prediction robust.

Fig. S1 shows 3-d visualizations of the epigenetic landscapes for small L (left column), respectively for a time-delay $\tau = 100$ (right column), see section S3 on inclusion of time-delay. These simulations can be compared to our standard parameter simulations in the middle column.

Main text Fig. 4 C shows how our model encompasses increases in gene copy number provided a sharp threshold for feedback activity. Fig. S2 A likewise shows that functional differentiation can be reproduced with a larger gene number, $N = 500$, when the sharper threshold for gene activity of the individual genes are parameterized by higher hill coefficient $h = 4$, and supplemented by a lower noise.

Fig. S2 B-G examines the success probability of simulations as function of r and μ at different hill coefficients, noise levels and nucleosome numbers. It is evident that the model is robust to changes in all parameters.

S3 - Pseudogenes, delayed feedback and early transient switching

Within the model settings that restrict the stable activation to a single OR gene, we considered the concept of pseudogenes. In this context these

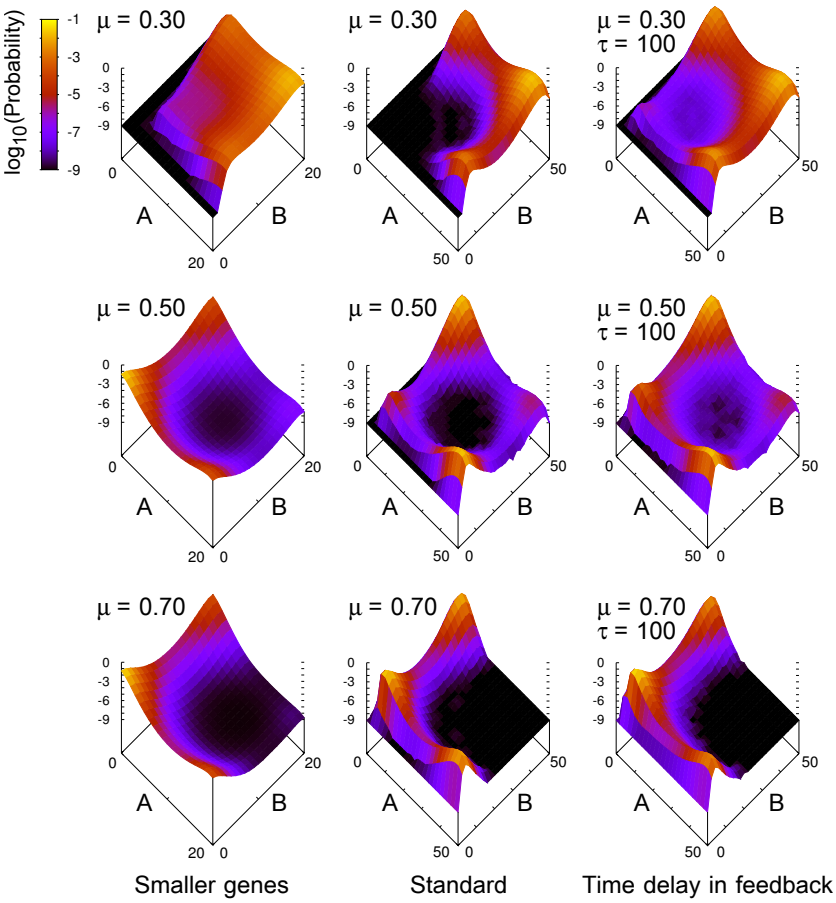

Fig. S 1: Related to Fig. 3 in main text. Epigenetic landscapes. Here we examine the probability (lighter colour for higher) for the two most active genes in the system, obtained by stochastic simulation over 10^8 time-units. The negative logarithm of this probability may be interpreted as an epigenetic landscape (62,63) with states that to varying degree prefer to be in the corners. The simulation uses standard parameters $N = 100$, $L = 50$, $r = 1$, $h = 2$, $\beta = 0.03$ and time delay $\tau = 0$ when nothing else is specified. Notice that both smaller L and a time delay lower barriers in the system, and favour transitions to the rightmost corner where two genes are active simultaneously.

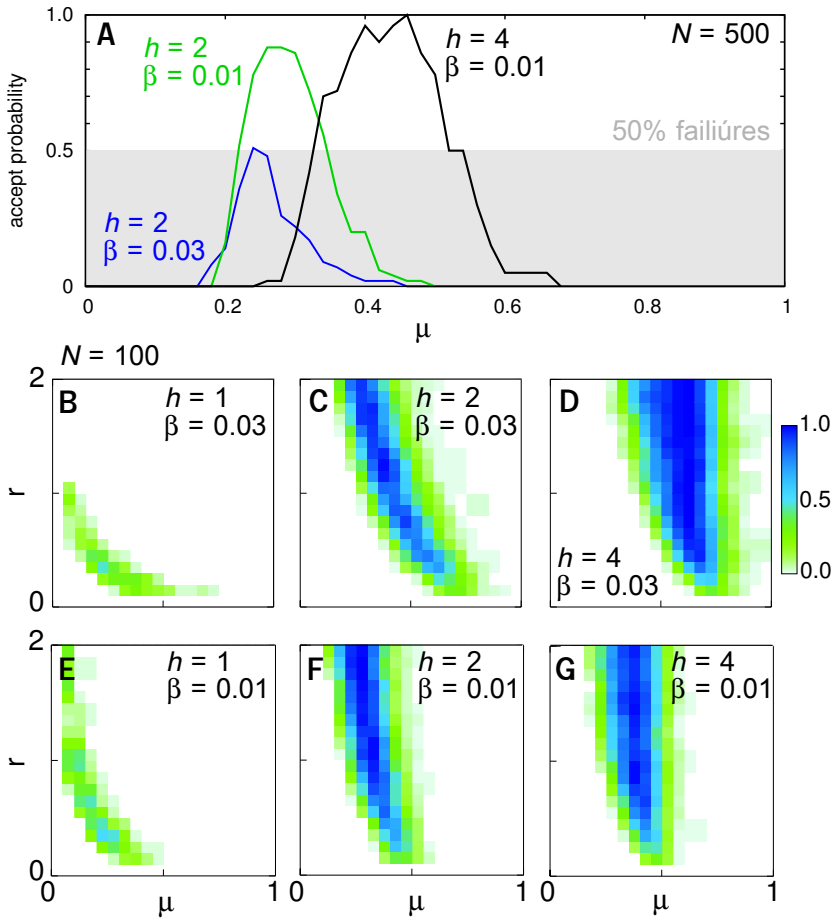

Fig. S 2: Related to Fig. 4 in main text. Robustness to changes in number of subsystems, noise and hill coefficient. **A** μ dependence of success probability for a system of $N = 500$ genes, with $L = 50$ nucleosomes and $r = 1$. Remaining parameters are as indicated in corresponding colours. **B-G** Success probability for a system of $N = 100$ genes as function of the values of r and μ calculated from 20 simulations. The simulations use standard parameters $L = 50$, $r = 1$ and time delay $\tau = 0$ with h and β as specified for each plot.

are subsystems affected by feedback as previously described but lacking the ability to produce the feedback. It will occasionally happen that first a pseudogene is activated, however, as the active pseudogene does not contribute to the negative feedback, R, another subsystem will eventually be activated and retain the dominant stable position, see Fig. S3 A.

In general developing neurons may show transient activation of different OR genes (24).Fixing the parameters at functional values like those of Fig. 3D in main text, initial transient activation of more OR genes is possible if we introduce a time delay, τ, between OR activation and feedback production. In this case we use

$$\frac{dP}{dt} = \frac{1}{\tau}(r \cdot \sum_j a_j^h - P) \tag{4}$$

with P then replacing $r \cdot \sum_j a_j^h$ in the expression for R. The effect is illustrated in Fig. S3 B.

Our standard model is appropriate when the factors facilitating the negative feedback have a life time that is shorter than it takes one gene to switch from silent to active state. In case the degradation is slower, the early activity of the OR system is often altered, with several genes turning on at very early times, for then subsequently to loose their activity when the full effect of the, then overproduced, negative feedback comes into play. As a result the qualitative behaviour from Fig. 4 in main text is reproduced with all time-delays that is substantially smaller than the maturation time of 1000 time-units, see supplementary Fig. S4.

The frequency of initial activation of a pseudogene is higher in systems without time delay due to the lack of feedback production from pseudogenes. In the case of time delay, real genes do not counteract their own activity while turning on and therefore switch as often as the pseudogenes, see Fig. S3 B.

Fig. S4 extends Fig. S3 by examining systematically the effect of time-delay τ on probability for successful differentiation. Furthermore, in the last panel of Fig. S4 we examine pseudogenes, again defined as olfactory genes that do not contribute to the global negative feedback. One sees that depending on an eventual time-delay pseudogenes do not always become activated with same probability as normal genes.

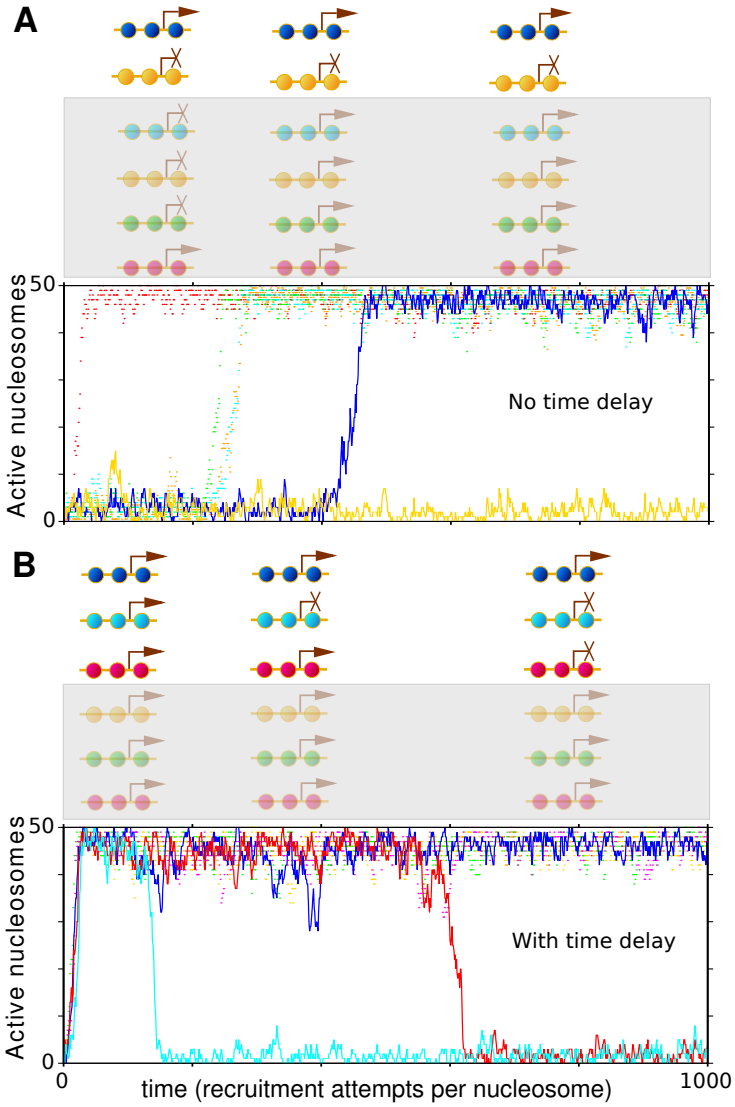

Fig. S 3: Pseudogenes and time-delay: **A** Activity of genes as function of time for our standard model with of $N = 100$ genes, and $N_{pseudo} = 40$ pseudogenes each covered by $L = 50$ nucleosomes, noise $\beta = 0.03$, asymmetry $\mu = 0.50$, overall repression factor $r = 1$, and hill $h = 2$. The dotted trajectories are pseudogenes, and illustrate that several pseudogenes may become activated early. The small promoter pictures show activity status of the correspondingly coloured genes in the covered time, with pseudogenes shown in gray shaded box. **B** As above, but with a time delay, τ, between OR activation and the feedback R. The time delay opens for transient activation of several OR genes. Once R accumulates the active ORs are repressed and only one OR gene remains active. As genes only sense their own activity after some time delay, the turn-on frequency for real genes is the same as for the pseudogenes. See also Fig. S3.

[h!]

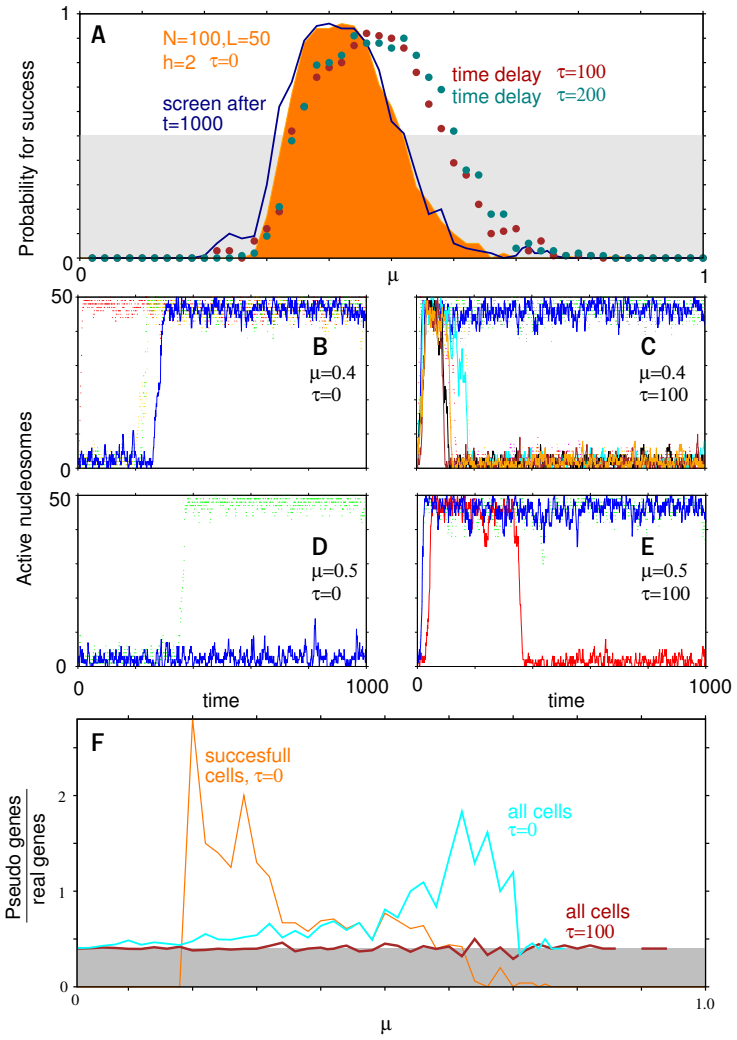

Fig. S 4: Related to Fig. 5 in main text. Investigation of time-delay for standard model with $N = 100$, $L = 50$, $\beta = 0.03$, $h = 2$, $r = 1$ and $N_{pseudo} = 40$ pseudogenes. **A** Success probability with and without time-delay. The orange area is standard conditions whereas blue curve is same simulation but only requiring that acceptance conditions are fulfilled after time 1000. The filled circles represent success probability with two different time delays τ. **B-E** Example of trajectories without (left) and with time-delay (right), and for 2 different values of bias μ. **F** Number of active pseudogenes in units of number of active normal genes. Pseudo-genes does not contribute to the negative feedback, and can more easily be turned on, cyan curve. Red curve show that with time delay, then pseudogenes are turned on exactly as much as real genes as indicated by gray area. Orange curve is same simulation as in cyan, but only counting turn-on frequency for cells that successfully turn on one and only one real gene in the simulation. Therefore measurement of pseudogene turn on rates will constrain our guess on μ and τ.

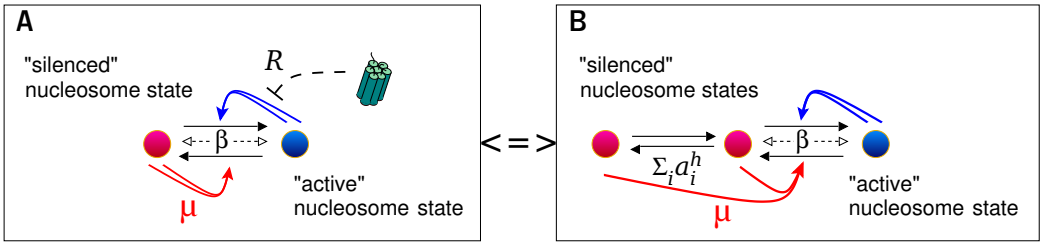

Fig. S 5: Form of OR neuron differentiation model. **A** In its general form the negative feedback, R, from the active OR genes simply reduces the chance for nucleosome activation. **B** In case the feedback acts through the binding and shielding of nucleosomes by a protein factor $P = r \sum a_j^h$. B exemplifies a specified version of the general case in A.

1. S4 - Forms of negative feedback

Figure S5 presents the general form of our model alongside with the outline of the model in the case where the negative feedback is envisioned as a binding and shielding protein factor, $P = r \sum a_j^h$.