Supplementary Appendix :

Mathematical model

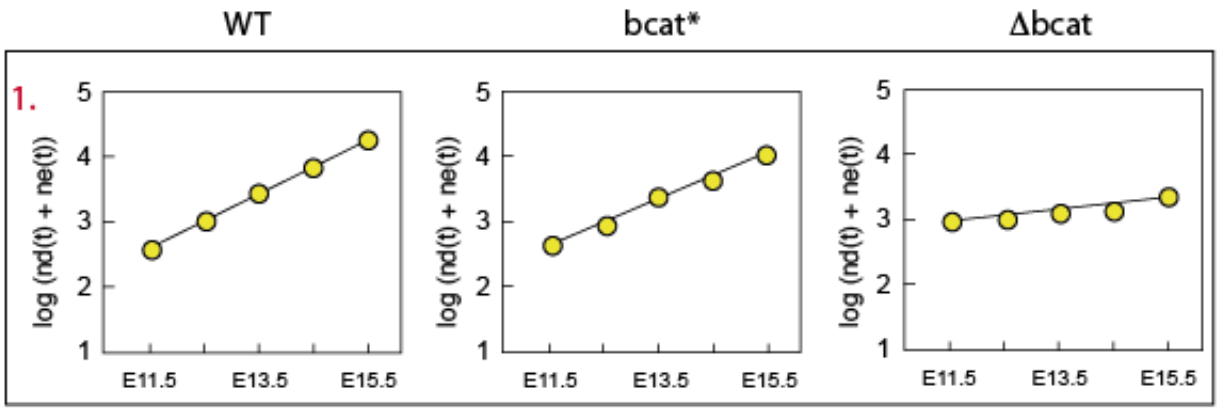
It should be pointed out that we first developed a linear model, which gave reproducible values for the doubling times of wt and bcat* melanoblasts. Unfortunately, this mathematical model was not reproducible and gave erratic results for Δbcat. We therefore decided to develop a generic mathematical model fitting any mouse mutant from white to black, with all possible intermediates. The mathematical model presented is non linear, which is biologically relevant. This non linear model generates reproducible values for the doubling times of wt, bcat* and Δbcat melanoblasts. It should be noted that the doubling time of wt melanoblasts was similar with both linear and non linear mathematical models. Similar findings were obtained for bcat* melanoblasts. The methodology developed here is a compromise between expected balance equations, behavior and feature extractions from data, and validation of data fitting.

A) Basic equations expressing the biological assumption. The number of founder melanoblasts located in the MSA is $n_{d,\theta}(t)$, which is equal to n^0. At this time, there are no melanoblasts in the developing epidermis $n_{e,\theta}(0) = 0$. The number of melanoblasts in the dermis (d) and epidermis (e) at a particular time (t) of development are denoted $n_{d,\theta}(t)$ and $n_{e,\theta}(t)$, respectively. The parameter θ represents the dependence on β-catenin activity. The flow $\Phi(t)$ of melanoblasts from the dermis to the epidermis is, unfortunately, unknown and cannot currently be determined experimentally. The dynamics of dermal and epidermal melanoblasts are modeled by the differential system

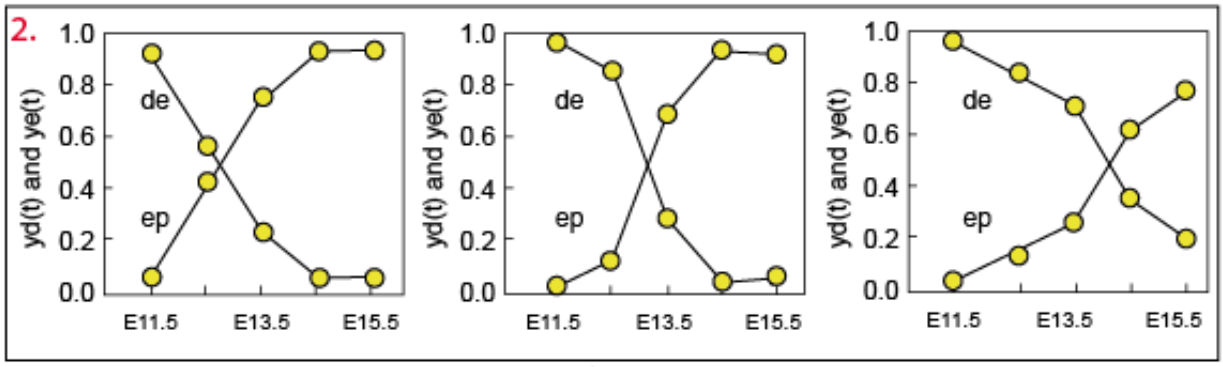
$$\frac{dn_{d,\theta}}{dt} = \mu_{d,\theta}(t)n_{d,\theta}(t) - \Phi_\theta\left(t, n_{d,\theta}(t), n_{e,\theta}(t)\right) \quad [1]$$ and

$\frac{dn_{e,\theta}}{dt} = \mu_{e,\theta}(t) n_{e,\theta}(t) + \Phi_\theta\left(t, n_{d,\theta}(t), n_{e,\theta}(t)\right)$ [2] , where $\mu_{d,\theta}(t)$ and $\mu_{e,\theta}(t)$ are the proliferation rates in the dermis and epidermis, respectively. The doubling times $\tau_{e,\theta}(t)$ and $\tau_{e,\theta}(t)$ are linked to $\mu_{d,\theta}(t)$ and $\mu_{e,\theta}(t)$ according to formula $\tau_{d,\theta}(t) = \frac{\log(2)}{\mu_{d,\theta}(t)}$, $\tau_{e,\theta}(t) = \frac{\log(2)}{\mu_{e,\theta}(t)}$ [3].

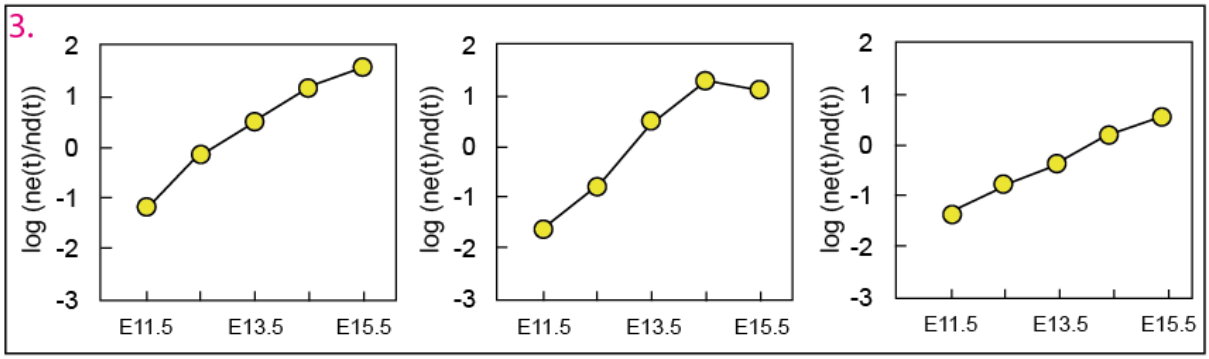
B) Additional knowledge extracted from the data. We need to determine the unknown flow $\Phi(t)$ to close the system. From the data, it is possible to find relevant features to guide the modeling process: **1.** Let us denote the total number of melanoblasts $n_\theta(t) = n_{d,\theta}(t) + n_{e,\theta}(t)$. Data analysis showed that the quantity $\log\left(n_\theta(t)\right)$ is an almost linear increasing function of time.



2. The fraction of melanoblasts in the dermis $y_\theta(t) = n_{d,\theta}(t)/n_\theta(t)$ was observed to behave as a decreasing sigmoid-shaped function of time. This allows us to postulate that $y_\theta(t)$ is the solution of a logistic-like differential equation $\frac{dy_\theta}{dt} = -\hat{c}_\theta(t) y_\theta(t)\left(1 - y_\theta(t)\right)$ [13].



3. It was also observed that the quantity $\log\left(n_{e,\theta}(t)/n_{d,\theta}(t)\right)$, obtained from biological data, was a quasi-linear increasing function of time.



C) Modeling of the flow term. From these observations and extracted features, the system [1] and [2] can be rewritten using the variables of interest $n_\theta(t)$ and $y_\theta(t)$. We get $\frac{dn_\theta(t)}{dt} = \mu_\theta(t)n_\theta(t)$ [14] and $\frac{dy_\theta(t)}{dt} = -\left(\mu_{d,\theta} - \mu_{e,\theta}\right)y_\theta(t)\left(1-y_\theta(t)\right) - \frac{\Phi_\theta(t)}{n_\theta(t)}$ [15] where $\mu_\theta(t) = \left(y_\theta(t)\right)\mu_{d,\theta}(t) + \left(1-y_\theta(t)\right)\mu_{e,\theta}(t)$. This can also be expressed as $\frac{d\log\left(n_{e,\theta}(t)/n_{d,\theta}(t)\right)}{dt} = \mu_{e,\theta}(t) - \mu_{d,\theta}(t) + \frac{\Phi_\theta(t)}{y_\theta(t)\left(1-y_\theta(t)\right)n_\theta(t)}$ [16]. As $\mu_{d,\theta}(t)$ and $\mu_{e,\theta}(t)$ are expected to be bounded and $d\log\left(n_{e,\theta}(t)/n_{d,\theta}(t)\right)/dt$ is observed to vary slowly, the last term in equation [16] must also be bounded. Thus, we decided to close the flux term $\Phi(t)$ as $\Phi_\theta(t) = \kappa_\theta(t)y_\theta(t)\left(1-y_\theta(t)\right)n_\theta(t)$ [17] where $\kappa_\theta(t)$ is a smooth bounded function, which is still unknown but expected to vary slowly in time. In this case, combining equation [15] with

the closure [17] gives [13] with $c_\theta(t) = \mu_{e,\theta}(t) - \mu_{d,\theta}(t) + \kappa_\theta(t)$. Equation $\frac{dy_\theta}{dt} = -c_\theta(t) y_\theta(t)\left(1 - y_\theta(t)\right)$ [13] is a logistic equation and thus is able to produce sigmoid-like solutions as observed in the data.

D) Estimation of *c* and *mu*. Finally, combining [17] with equation [16] gives $c_\theta(t) = d\log\left(n_{e,\theta}(t)/n_{d,\theta}(t)\right)\Big/dt$ [7]. Equation [1] can be exploited to estimate the function $c_\theta(t)$ from the data. We will denote its estimator $\hat{c}_\theta(t)$. Similarly, equation [14] can be rewritten $d\log\left(n_\theta(t)\right)\Big/dt = \mu_\theta(t)$ [6]. Numerical experiments show that both $\mu(t)$ and $c(t)$ are smooth functions of time. Correct fitting to the data fitting requires an accurate estimation of these functions. Typically, the ratio between the maximum and the minimum of these functions is 10.

E) Final system used to estimate the number of founder melanoblasts. Note that the system can be completely closed without any dependence on the unknown function $\kappa_\theta(t)$:

$$\frac{dn_{d,\theta}(t)}{dt} = \hat{\mu}_\theta(t) n_{d,\theta}(t) - \hat{c}_\theta(t)\frac{n_{d,\theta}(t) n_{e,\theta}(t)}{n_{d,\theta}(t) + n_{e,\theta}(t)} \quad [18]$$

and

$$\frac{dn_{e,\theta}(t)}{dt} = \hat{\mu}_\theta(t) n_{e,\theta}(t) - \hat{c}_\theta(t)\frac{n_{d,\theta}(t) n_{e,\theta}(t)}{n_{d,\theta}(t) + n_{e,\theta}(t)} \quad [19]$$

. Once the functions $c_\theta(t)$ and $\mu_\theta(t)$ are estimated, it appears that the system [18] & [19] is able to give solutions that fit the data almost perfectly using standard estimation algorithms. To evaluate the number of founder melanoblasts, we initially interpolated the slope as is to earlier time from the WT curve. The values obtained were between 15 and 25. To avoid any stringency, we applied an inferior bound of 10 and superior bound of 50. In the second round, the number of founder melanoblasts was estimated by solving an inverse problem using least square minimization

given by $\min_{n_0,\, 10 \le n_0 \le 50} \sum_{i=E10.5}^{E15.5} \frac{1}{2\sigma_i^2}\left(n^{meas,i} - n^i(n_0)\right)^2$. The standard deviations σ_i are weighting variables that takes into account the errors associated with the data. For each mouse type, we computed an initial fraction y. We found $y^0 = 1 - 810^{-4}$. Obviously, " 8 10^{-4} " has no biological reality because at E8.5, at early developmental stage, stochastic and discrete dynamics are predominant. This value is, however, important to allow correct fitting. Regarding the number of founders, the estimation process returns a value of 16 ± 0.1 for each mouse type. This means that on one side of the embryos there is roughly one founder melanoblast for about two somites. The uncertainty about the value is related to the model uncertainty, which unfortunately cannot be quantified. The uncertainty is undoubtedly relatively large because a continuous differential model is used to represent the stochastic discrete dynamics in this early development stage.

F) Compatibility relationship of proliferation rate and the unknown kappa function. Equations [7] and [6] can be used to estimate both the function $c_\theta(t)$ and $\mu_\theta(t)$ from the data. We denote their estimators $\hat{c}_\theta(t)$ and $\hat{\mu}_\theta(t)$. Once the estimators $\hat{c}_\theta(t)$ and $\hat{\mu}_\theta(t)$ are computed using [7] & [6], we obtain the algebraic compatibility equations $\hat{c}_\theta(t) = \kappa_\theta(t) + \mu_{e,\theta}(t) - \mu_{d,\theta}(t)$ [4] and $\hat{\mu}_\theta(t) = \left(y_\theta(t)\right)\mu_{d,\theta}(t) + \left(1 - y_\theta(t)\right)\mu_{e,\theta}(t)$ [5]. Another result is that the determination of doubling times in the dermis and epidermis is an ill-posed problem from a deterministic point of view. At a given time, the system of algebraic equations [4] and [5] is underdetermined, with two equations and three unknowns $\mu_{d,\theta}(t)$, $\mu_{e,\theta}(t)$ and $\kappa_\theta(t)$. If $\kappa_\theta(t)$ can be estimated, then $\mu_{d,\theta}(t)$ and $\mu_{e,\theta}(t)$ can be determined. Rearranging the terms into [4] and [5] gives $\mu_{d,\theta}(t) = \hat{\mu}_\theta(t) - \left(1 - y_\theta(t)\right)\left(\hat{c}_\theta(t) - \kappa_\theta(t)\right)$ [9] and $\mu_{e,\theta}(t) = \hat{\mu}_\theta(t) + y_\theta(t)\left(\hat{c}_\theta(t) - \kappa_\theta(t)\right)$ [8].

G) Estimation of the doubling time according to relevant biological constraints. A rough estimation of $\kappa_\theta(t)$ can be obtained by adding a priori knowledge about doubling times. A priori biological knowledge about doubling times, in particular $\tau_{e,\theta}(t) \le \tau_{d,\theta}(t)$ and, $\tau_{d,\theta}(t) \le 3\tau_{e,\theta}(t)$ allows us to set bounds for the random variable $\kappa_\theta(t)$. Under the assumption $\tau_{e,\theta}(t) \le \tau_{d,\theta}(t)$, we have $\mu_{d,\theta}(t) \le \hat{\mu}_\theta(t) \le \mu_{e,\theta}(t)$ and then from [9] & [8] we get the inequality $\kappa_\theta(t) \le \hat{c}(t)$. A second a priori constraint gives the lower bound on $\kappa_\theta(t)$; $\kappa_\theta(t) \ge \left(\hat{c}_\theta(t) - 2\mu_\theta(t)/3 - 3y_\theta(t)\right)$.