Impact of Zygosity on Bimodal Phenotype Distributions

T. Holst-Hansen, E. Abad, A. Muntasell, M. López-Botet, M. H. Jensen, A. Trusina, J. Garcia-Ojalvo

Abstract

Allele number, or zygosity, is a clear determinant of gene expression in diploid cells. But the relationship between the number of copies of a gene and its expression can be hard to anticipate, especially when the gene in question is embedded in a regulatory circuit that contains feedbacks. Here we study this question making use of the natural genetic variability of human populations, which allows us to compare the expression profiles of a receptor protein in natural killer cells between donors infected with human cytomegalovirus (HCMV) with one or two copies of the allele. Crucially, the distribution of gene expression in many of the donors is bimodal, indicative of the presence of a positive feedback somewhere in the regulatory environment of the gene. Three separate gene-circuit models differing in the location of the positive feedback with respect to the gene can all reproduce the homozygous data. However, when the resulting fitted models are applied to the hemizygous donors, one model (the one with the positive feedback located at the level of gene transcription) is superior in describing the experimentally observed gene-expression profile. In that way, our work shows that zygosity can help us relate structure and function of gene regulatory networks.

Author Summary

Nearly all mammalian cells, including human cells, have two copies of each chromosome, and thus possess two potentially different copies of each gene (which might be in some cases non-functional or even absent). Naïvely one might expect that two identical copies of the gene would lead to the protein being expressed at twice the rate, but many factors can alter this simple calculation. One of these factors is the existence of feedback mechanisms affecting in one way or another the regulatory circuit in which our gene of interest is embedded. Here we study the relationship between the number of gene copies and the expression of a receptor protein that plays a crucial role in the recognition of pathogens by natural killer cells, which are important elements of the innate immune system. Experimental data of virus-infected donors reveals a bimodal expression profile of this receptor, typical of a positive feedback, and a clear difference between donors with one or two copies of the gene. Mathematical modeling allows us to find the likely location of the feedback loop within the gene's regulatory circuit, by requiring the correct model to reproduce the expression profiles of both types of donors.

Introduction

Natural killer (NK) cells are part of the innate immune system. They recognize potential threats through a gauging process, in which signals from activating and inhibitory receptors are integrated into a decision. A target cell is killed either if inhibitory receptors are not engaged, or if the NK cell receives a strong activating stimulus. Inhibitory receptors sense MHC class-I surface molecules that present intracellular

peptides [1]. Since MHC molecules can be lost due to mutations in tumors or viral infection, NK cells are particularly important in the defence against those threats [2, 3]. The innate immune system thus provides a generic response to pathogens, and contrary to the adaptive immune system, it does not create long-term immunity. However, NK cells exhibit adaptive memory features in response to certain viruses [4]. For instance, in response to infection with human cytomegalovirus (HCMV), some donors experience stable expansion of the NK subset containing the activating NKG2C receptor [5–7], resembling the memory of the adaptive immune system. The NKG2C+ NK cell subset is believed to play a role in antiviral defense, and has been shown *in vitro* to mediate a potent antibody-dependent response against HCMV-infected cells [8].

Notably, HCMV-infected donors usually expand the NKG2C+ NK cell subset in a bimodal fashion [9], displaying two distinct phenotypes with either high ('bright') or low ('dim') receptor expression. These two phenotypes differ in their modulation of key immune cell functions such as degranulation, proliferation, and sensitivity to interleukin stimulation [9]. Interestingly, the distribution of NKG2C expression is affected by the number of copies (zygosity) of the gene encoding the membrane receptor NKG2C. Specifically, homozygous (two gene copies) and hemizygous (one gene copy) donors differ significantly in the expression level and function of the NKG2C receptor, as well as in the fraction of NK cells that express this receptor [9]. The effect of zygosity on HCMV-induced bimodality thus provides us with a strong constraint that could shed light on how the virus interacts with the immune system.

Bimodal distributions may be produced by a bistable response, which is commonly generated by positive feedbacks [10–13]. In this paper we show that three types of positive feedback, located at different positions with respect to the NKG2C gene (upstream of the gene transcription, directly at the level of gene transcription, and at the post-transcriptional level) can generate similar bimodal distributions in homozygous donors. However, a comparison with the corresponding hemizygous distributions allows us to differentiate between these three cases. Specifically, our computational model shows that the distributions of NKG2C+ NK cells are best described by a positive feedback at the level of NKG2C transcription, as opposed to positive feedbacks located pre- or post-transcriptionally. We use bifurcation analysis to show mathematically that zygosity leads to two types of changes in the bistable expression distributions: changes in expression level, and changes in the fraction of cells expressing the phenotype. As we will see, a positive feedback upstream of the gene transcription leads only to expression level changes, while a post-transcriptional positive feedback just leads to changes in the fraction of expressing cells. In contrast, a positive feedback acting directly at the level of gene transcription has both effects in response to zygosity changes: in the expression level and in the fraction of expressing cells.

Copy number variations (CNVs) are a common form of genetic variability, and are linked to various autoimmune diseases such as rheumatoid arthritis [14], asthma [15], and susceptibility towards HIV [16]. Few quantitative studies, however, have addressed so far the influence of copy number on gene circuits. In particular, circuits containing positive autoregulation have been studied in the contexts of sex determination in fruit flies [17] and Down's syndrome in mice and humans [18]. Also, the effect of copy number on gene circuit dynamics at the single-cell level has been examined in bacterial genetic competence [19]. But in spite of these efforts, the link between CNVs and the response of gene regulatory circuits is still elusive, in particular in human cells. The mathematical framework presented here can serve as a tool for using genotypic information in combination with molecular biology assays, to better understand the connection between the topological architecture and the function of cellular regulatory networks.

Methods

Fitting and Model Comparison

Fit quality is quantified by Pearson's χ^2-value, which together with the number of degrees of freedom (ndf) provides a measure of the quality of the fit. We also use the Akaike information criterion (AIC) as a measure of relative model quality. The AIC value is given as: $\mathrm{AIC} = 2k - 2\ln(L)$, where k is the number of estimated parameters and L is the maximum value of the likelihood function. Assuming that residuals are distributed according to identical independent normal distributions, the Akaike Criterion can be rewritten to $\mathrm{AIC} = 2k - n\ln(\mathrm{RSS}/n)$, where RSS is the residual sum of squares $\sum_{i=0}^{n}\left(y_i - f(x_i)\right)^2$. For small sample sizes the AIC is biased, but can be corrected by the addition of an extra term: $\mathrm{AIC}_c = \mathrm{AIC} + \frac{2k(k+1)}{n-k-1}$ [20]. We calculate a combined Akaike score for each model using the results for both types of zygosity. The relative strength of evidence for each model is proportional to $\exp\left(-\frac{1}{2}\Delta_i\right)$, where $\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min}$. This can be summarized by a set of weights:

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{k=1}^{N}\exp\left(-\frac{1}{2}\Delta_k\right)}$$

where $\sum_{i=1}^{N} w_i = 1$. These weights determine the model that best represent the data relatively to the others [21].

Results

Effect of zygosity on NKG2C expression distributions: experimental observations

In previous work [9], NK receptor expression profiles of fresh peripheral blood samples of healthy human donors were measured by flow cytometry (see Ref. [9] for details of the experimental protocols). In the present study we focus on the distribution of NKG2C expression, which clearly shows that many HCMV-positive donors have bimodal profiles (Fig 1A,B). Data from nullzygous donors (homozygous for the NKG2C deletion, Fig 1C) shows that the left peak of the bimodal distribution is purely background fluorescence, since these individuals do not express the receptor. Based on the nullzygous donor profiles we define an expression threshold of $R = \log_{10}(\mathrm{NKG2C}) = 2.75$ (leftmost vertical dashed line in Fig 1A-C). Cells are termed 'expressing' if they have an expression level above that threshold (Fig 1D). The fraction of cells that express NKG2C is calculated as the area under the normalized NKG2C distribution beyond the expression threshold. Similarly, the average expression level is only calculated for cells above the expression threshold. The data reveals a significant decrease in both the expression level (Fig 1E) and fraction (Fig 1F) of NKG2C+ NK cells, when comparing homozygous with hemizygous donors. In the two cases the distribution of NKG2C in the NK population of individual donors is bimodal, but also highly variable between donors.

A key assumption in the comparison of the data with the models described below is that the fluorescence measurements are proportional to the surface level of NKG2C, which is a standard and reasonable assumption in flow cytometry [22]. We also assume that the NK population is in a state of long-lived homeostasis, as has been shown experimentally [23]. In particular, NKG2C+ NK cells are known to remain at stable steady-state levels for years in healthy donors [9, 24].

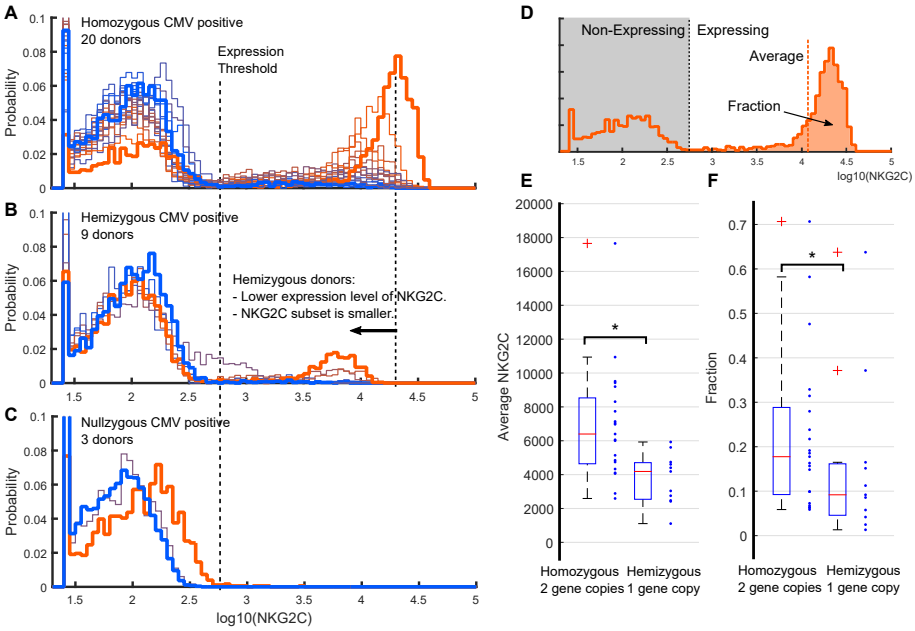

Fig 1. Zygosity influences expression level and prevalence of NKG2C in NK cells of HCMV positive donors. Flow cytometry measurements of NKG2C in NK cells [9], for HCMV positive homozygous (**A**), hemizygous (**B**), and nullzygous (**C**) donors. Individual donors are color-coded according to the NK population fraction expressing the NKG2C receptor, with orange (blue) corresponding to the highest (lowest) expressed fraction. Panel **D** shows how the distributions are quantified. Specifically, the fraction of expressing cells is defined as the area under the normalized distribution above the expression threshold. Likewise, the average expression level of cells is calculated for those cells above the expression threshold. Panels **A** and **B** show that reducing gene copy number lowers both the expression level and the fraction of NK cells expressing NKG2C. Panels **E**, **F** show box plots quantifying the significant differences between homozygous and hemizygous donors in the average NKG2C level of expressing NK cells, and in the fraction of the NK population expressing NKG2C. In those two panels, the asterisk indicates a statistically significant difference, using a two-sided Mann-Whitney U test with significance level of $p = 0.05$ (see Supporting Table 1).

Modeling NKG2C bimodality

The aim of the model is to describe the impact of zygosity (gene copy number) on the expression level of the NKG2C receptor in NK cells, and on the fraction of cells in the population expressing this receptor. A key observation that constrains the model is that the population is bimodal in its expression of NKG2C. We assume that this bimodality is caused by a positive feedback that induces a bistability at the single-cell level. It is important to note that bimodality arises whenever the receptor distribution expresses two primary modes, while bistability corresponds to the coexistence of two stable equilibrium states. Bimodality does not necessarily require bistability in the system, but bistability generally leads to a bimodal response [25]. Thus, for the sake of simplicity we assume here that the bimodality observed in this system is caused by bistability.

We consider three models that differ on where the positive feedback is located with respect of NKG2C expression (Fig 2A). In model **A** the feedback occurs upstream of NKG2C expression, at the level of a transcription factor regulating the expression of the gene, named pre-transcriptional in what follows. In model **B** it arises at the level of transcription of the gene. Finally, in model **C** the feedback is considered to occur at the post-transcriptional level. All three models describe the dynamics of three variables: a transcription factor T, the mRNA m, and the mature receptor R. The model is kept minimal, in order to illustrate the effects of bistability and gene copy number, and as

such it does not take into account specific proteins or processes. The variables could therefore correspond to any part of the signalling pathway with a positive feedback onto itself.

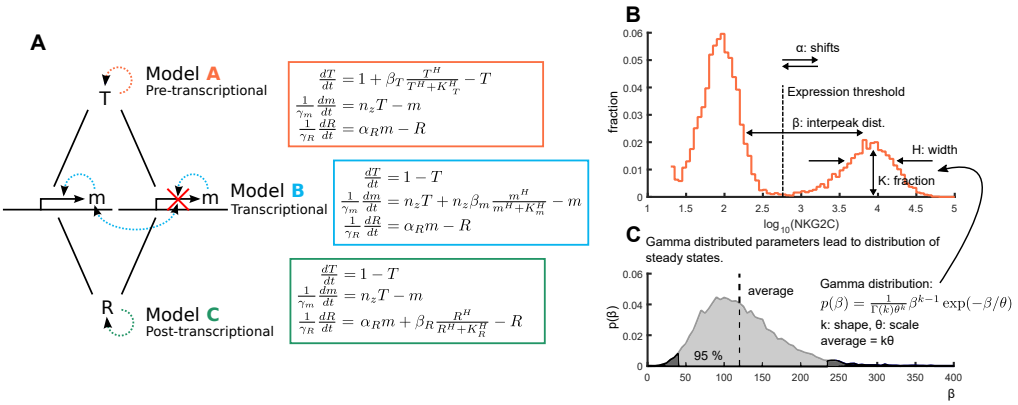

Fig 2. Three minimal models of NKG2C expression. **A**: The three models differ in the position of the feedback relative to transcription; A, a pre-transcriptional, B, a transcriptional and C, a post-transcriptional positive feedback. Each model is described by three differential equations governing the dynamics of T, a transcription factor, m, mRNA concentration, and R, mature receptor. The parameter n_z denotes the gene copy number in each model, and the positive feedback is incorporated as a Hill term. **B**: Each model parameter affects the shape of the overall distribution in different ways: α shifts the entire distribution (only model A and B), β changes the distance between the two peaks, K controls the fraction of cells in each peak and H the sharpness of transition between peaks, i.e. the width of the peak and especially the region between the two peaks. **C**: The distribution of the NKG2C receptor (R) is generated by drawing the parameters α, β, K and H from gamma distributions.

In model **A**, for instance, T has a positive feedback onto itself implemented with a Hill term. For simplicity we describe production and degradation as linear terms:

$$\begin{aligned}\frac{dT}{dt} &= \alpha_T + \beta_T \frac{T^H}{T^H + K_T^H} - \gamma_T T \\ \frac{dm}{dt} &= n_z \alpha_m T - \gamma_m m \\ \frac{dR}{dt} &= \alpha_R m - \gamma_R R\end{aligned} \tag{1}$$

Here α_T is a basal production, α_m and α_R are linear production rates, β_T is the strength of the positive feedback, H is the Hill coefficient, K_T is the activation threshold and γ_x are degradation rates. n_z is an integer that corresponds to gene copy number, and which will be varied in what follows to account for the differences between homozygous and heterozygous donors. Note that the model describes the receptor level, while the donor distributions are measures of fluorescence from flow cytometry. This means that the first peak in the experimental measurements actually corresponds to background activity in the absence of receptors. In the model we reproduce this activity through the basal production coefficient α_T. Each model has nine parameters, but by rescaling we reduce the number of variables to six (see Supporting Information). Models **B** and **C** are similar to model **A** above, with the positive feedback on different variables. The rescaled equations are shown in the right panels of Fig 2A. We have not scaled R in terms of α_R, as we want to maintain that parameter as a fitting parameter.

We assume that the NK population of donors is in homeostasis at the time of measurement, with no sustained dynamics. Therefore we only consider steady states,

which for the rescaled version of model **A** are given by the following equations:

$$\begin{aligned} 0 &= 1 + \beta_T \frac{T_s^H}{T_s^H + K_T^H} - T_s \\ m_s &= n_z T_s \\ R_s &= \alpha_R m_s \end{aligned} \tag{2}$$

For $T \gg K_T$, the first equation reduces to $T_s = 1 + \beta_T$, while in the regime where $T \ll K_T$, $T_s \approx 1$ if $\beta_T / K_T^H \ll 1$, which is a good approximation in our case because the majority of cells are not expressing the receptor. This approximate solution shows that the separation between peaks is controlled by the β parameter (Fig 2B). The Hill coefficient H, in turn, controls the width of the peaks, since a higher coefficient leads to a sharper transition between solutions. Finally, the threshold K_T of the positive feedback regulates the fraction of cells expressing the receptor, and α scales both solutions simultaneously. The same is true for model **B**, while in model **C**, α only controls the position of the first peak. Note that the steady states also predict no receptor production in the case of nullzygous donors.

Noise is necessary to generate a population of diverse cells. We assume that differentiated NKG2C+ NK cells do not switch receptor expression during their life-span, so that the bimodality occurs only at the population level. Therefore we have chosen to introduce noise by drawing parameters from distributions, rather than adding dynamical noise via stochastic simulations, as is commonly done when studying stochastic gene expression. The abstract character of our model also makes stochastic simulations less suitable, since intermediate steps that could introduce noise are not made explicit. Each parameter is drawn from a gamma distribution (Fig 2C), described by a shape coefficient k and a scale coefficient θ, so that the average of the distribution is $k\theta$. The gamma distribution is a good description of the statistics of protein abundance [26]. In the Supporting Information we show how noise in each parameter affects the distribution. The distribution of NKG2C expression in the NK population is simulated by drawing a set of parameters (α, β, K, and H) for each cell. Each of the four parameters defining the steady state (Eq 2) is drawn from a distribution with given average and scale coefficient, which gives eight free parameters in total. The NKG2C distributions were generated by numerically solving the nonlinear equations (2) for the steady states of model **A** (and the corresponding steady-state equations for the other models). In those cases where there were two stable solutions we chose the lowest one. This choice can be justified because in the 4-d space where the parameters are chosen from, the bistable region is relatively narrow compared with the full parameter distribution, and thus the difference between choosing the low and high solution in the bistable region is relatively small. We also note that, as can be seen in Supporting Fig S4, generating the receptor distributions from the steady-state equations leads to results that are similar to solving numerically the differential equations given in Fig 2A and letting the system equilibrate. This indicates that the relaxational behavior of the system can be neglected.

Transcriptional Positive Feedback Best Captures Zygosity Impact in HCMV Positive Donors

To compare each of our models with the experimentally measured NKG2C expression levels of the homo- and hemizygous HCMV positive donors (Fig 1), we calculated a profile of the "median donor" for each group, as the median frequency of the cell number at each expression level. We then fitted each model to the homozygous donor group and achieved fits of similar quality (measured in terms of the χ^2-value), as shown in Fig 3A (see Methods section and Supporting Information for fitting procedure). Hence, it is not possible to differentiate between the three models based only on homozygous donors.

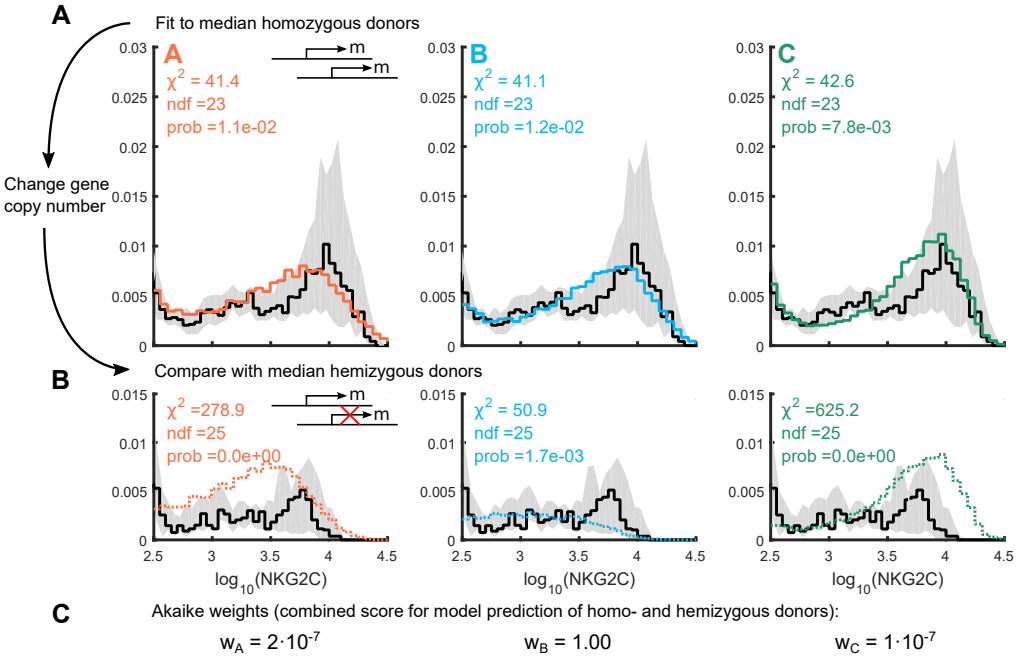

Fig 3. Model B makes the best prediction of NKG2C expression in hemizygous donors. **A**: Each model is fitted to a median homozygous donor (black lines), generated by taking the median cell frequency at each receptor level. The grey area shows the 50% interquartile region. The model distributions provide fits of similar quality (color lines) based on a Pearson's χ^2-value. **B**: Starting from the fit of the median homozygous donor, the gene copy number, n_z, is changed from two to one, and the similarity between the median hemizygous donor and the models are scored with a χ^2-value. **C**: The Akaike weights are calculated based on both the homo- and hemizygous simulation, and are indicative of the quality of each model in fitting the data, compared with the other models. Model B is clearly favoured over both A and C. All distributions were generated from 100.000 simulated cells.

Using the parameters found by fitting the median homozygous donor, we next changed the gene copy number n_z from two to one. Each model makes different predictions of the median NKG2C distribution of hemizygous donors, which makes it possible to distinguish between models (Fig 3B). In particular, model **B** provides the best prediction of the hemizygous donor group, indicating that a positive feedback at the transcriptional level is the most likely of the three possibilities.

The models we present here are simple and do not capture all details of the distributions. Thus, as an alternative measure of relative model quality to the χ^2-value, we also used the Akaike information criterion (AIC). The Akaike weights, shown in Fig 3C, are based on the combined model prediction of homo- and hemizygous donors. Model **B** is again the most likely of the three models in describing the NKG2C changes due to zygosity. Biologically, this implies that upstream or downstream positive feedbacks alone are not sufficient to describe, at the single cell level, the impact of zygosity on NKG2C expression.

From a modeling point of view the virus has the effect of increasing the strength of the positive feedback, β, causing higher levels of NKG2C in the NK population. It also lowers the activation threshold K of the positive feedback, which increases the fractional expression of NKG2C. This could occur for instance through promoter alterations or increased receptor activation. But our analysis shows that if the bimodality is caused by a positive feedback, it is most likely to start and end at the transcriptional level, as described by model **B**.

Bifurcation Diagrams Present Model Differences

We can use bifurcation diagrams to translate a distribution of parameters into a distribution of NKG2C expression. Changing the model zygosity will alter the bifurcation diagram, and therefore the corresponding receptor expression. In Fig 4 we show the bifurcation diagram of each model for varying activation thresholds K, and depict how a distribution of K is translated into receptor expression. Our analysis shows that all three models undergo two saddle-node bifurcations, delimiting the region in which three stable fixed points exist (two of them stable, represented by solid lines, and one unstable, depicted by a dashed line). For low activation thresholds a single stable solution exists

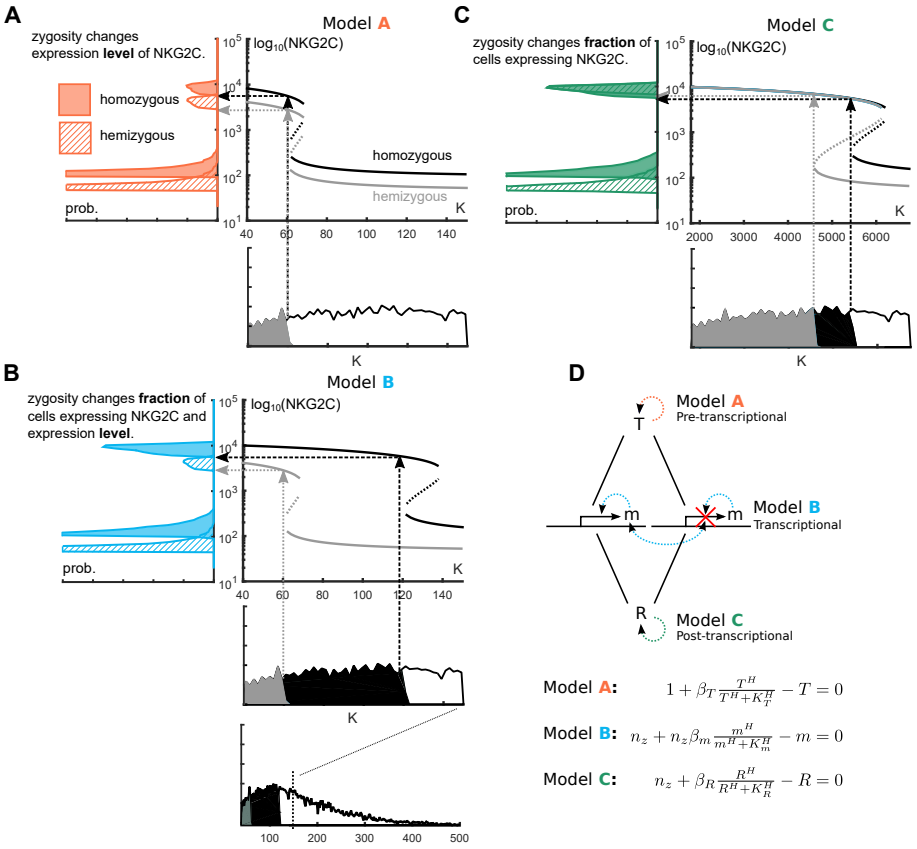

Fig 4. Quantitative model differences are explained by how zygosity influences bifurcations. NKG2C distributions as produced by noise in the K parameter. Each K parameter is translated through the bifurcation diagram into a receptor level. **A**: Model A (pre-transcriptional positive feedback): Zygosity does not shift position of the bifurcations in K. The solution values are however shifted by a factor two. **B**: Model B (transcriptional positive feedback): Zygosity changes both position of bifurcations and solution level. This translates into a change in both level and fraction of NK cells expressing NKG2C. **C**: Model C (post-transcriptional positive feedback): The bifurcation is shifted by a zygosity change while the high value solution falls on top of each other, i.e. the fraction is changed but not the receptor level. The K parameter is drawn from a gamma distribution with $\mu(K) = 160$ and shape $\theta(K) = 60$. **D**: Sketch of the regulatory network in each model. The nonlinear steady state equation is affected differently by n_z in each model, leading to the different bifurcation diagrams.

at high NKG2C levels. As the activation threshold increases, the low stable solution and an unstable solution is created. Further increase of K causes the unstable and the high stable solution to annihilate. The position of the bifurcation determines the fraction of the K-distribution that is translated into expressing cells, while the value of the solution corresponds to the NKG2C expression level. We adjust the position and width of the parameter distributions relative to the bifurcations so as to recover as closely as possible

the experimentally observed receptor distributions.

In model **A** the bifurcations are positioned at the same K-values independent of zygosity, while the solutions are shifted by approximately a factor two. This results in a change in expression level, but not in the fraction of cells expressing the receptor. In contrast, in model **C** zygosity causes a shift in the first bifurcation while the other remains the same. This lead to a changed fraction of expressing cells that is too small to reproduce the experimental observations. The effect of zygosity on the high stable solution is negligible, which causes the position of the second peak to remain approximately constant. Finally, in model **B** zygosity shifts both the location of the two bifurcations and the level of the solutions, leading to noticeable changes in both expression level and the fraction of cells expressing the receptor. Similar bifurcation diagrams can be made for the β and H parameters, see supporting information.

Mathematically the source of bistability in all models is the nonlinear term. In the expression of the steady state solution of each model we can see the relation between copy number and the nonlinearity (Fig 4D). The gene copy number, n_z, does not directly impact the nonlinear equation of model **A**, which explains the constant fraction of expressed cells exhibited by this model. The solutions will be multiplied by n_z, which causes the shift in the expression level. In model **B**, in contrast, n_z affects both the nonlinearity (and thus the fraction of expressing cells) and the solution level (and thus the expression). Finally, in the nonlinear equation of model **C** the gene copy number only changes the basal production term. This changes the fraction of expressing cells but has a small impact on the high receptor solution, because β_R is large compared with n_z.

In conclusion, model **B** is the only model that predicts both changes in expression level and fraction with zygosity, as observed among HCMV positive donors in Fig 1. The bifurcation diagrams further show that the location of the positive feedback relative to transcription gives three qualitatively different types of behaviour. Distributional changes due to zygosity can therefore help locate positive feedbacks in bimodal distributions.

Discussion

The role of copy number variation in disease susceptibility and pathogenesis is becoming increasingly more recognized [27, 28]. However, the relation between gene copy number and phenotype expression is only partially understood. Inspired by an observed phenotypic difference between homo- and hemizygous HCMV positive donors in natural killer cells [9], we have shown that in the special case of a bimodal phenotype caused by a bistability, the expression differences due to zygosity can provide valuable information.

Model validity and fitting

Within each donor group (zygosity and seropositivity) there were large variations in expression, ranging from no expansion to strong bimodal expression of the NKG2C+ NK cell subset. To test if the use of a median donor was reasonable under these conditions we did a cluster analysis (see Supporting Information). The analysis showed that seropositive donors can be divided into two subgroups: no expansion and bimodal distribution. The cluster analysis further showed that for donors with bimodal expression, the position of the second peak is a strong biomarker of zygosity. We have not shown any fits of single donors, but each model fits a large range of distributions reasonably well (see Supporting Information). It is also worth noting that the hemizygous distributions are not directly fitted to the models, but are compared with the fitting obtained from the homozygous donors (after accounting for the change in zygosity). This is a more stringent comparison between the models than performing different fits in the two cases.

If we assume that not all cells in homozygous donors have two functioning genes, the model can be modified by introducing a subpopulation which effectively has one gene. This would create a superposition of a homo- and hemizygous distributions, which could create better fits, but also introduces an extra free parameter. Notice this does not change the results from the bifurcation diagrams, but zygosity effects will decrease as the fraction of cells with one functioning gene increases.

A more detailed regulatory network of the biochemical processes involved in receptor expression could be interesting in terms of identifying positive feedbacks. However, for the purpose of this study a complex model would distract from the point of our three conceptual models.

Biological interpretation of positive feedbacks

Our modeling study suggests that, in order to reproduce the bimodality observed experimentally, the feedback has to include a transcriptional component. All the models considered here can be interpreted as single-cell models, but the processes are not necessarily restricted to individual cells. For instance, a bistable T-variable could be an upstream transcription factor, but also a bistable external input. The T-variable can therefore be any or all processes upstream of transcription.

Model **C** is more restricted, as a post-transcriptional positive feedback can only be between transcription and the mature receptor in the cell membrane. Even though R is set equal to the membrane protein in this study, the positive feedback can be located downstream of transcription but upstream of the mature receptor. The effect would be the same if the positive feedback is followed by linear processes. A candidate for this type of behaviour could be the CD94 protein, which dimerizes with NKG2C to create the mature receptor [29]. Any bistability in CD94 would be reflected in the phenotype expression.

In model **B**, in turn, the positive feedback could in principle extend outside the cell as long as its impact reaches the production of mRNA in the end. It has been suggested that NKG2C and the inhibitory receptor NKG2A, which both dimerize with CD94, are mutually exclusive in CD8+ T-cells [30]. Another study has shown that NKG2C and NKG2A were reciprocally expressed in CD56dim NK cells, but co-expressed in CD56bright NK cells [31]. Receptors which mutually inhibit each others' transcription is an example of a positive feedback at the transcriptional level.

Positive feedback can also arise at the population level. Previous studies [9] have shown that there is a correlation between proliferation rate and NKG2C level. This could lead, assuming that daughter cells inherit their mother's receptor levels, to a positive feedback at a population level, since NKG2C-expressing NK cells will grow faster than other NK subsets. Adding this effect would require, however, a modeling approach qualitatively different from the one considered here. NK proliferation has previously been modeled [32], but not in a homeostatic state and not in relation to the NKG2C receptor. Proliferation-dependent bimodality would thus deserve further investigation.

Zygosity and gene copy number effects

Zygosity has previously been identified as a source of NK receptor alterations, but through its effect on the receptor ligand rather than on the receptor itself. Specifically, it has been shown that the zygosity of HLA-Cw7 (coding for NK receptor ligand) altered the NK CD158+ subset [33]. These alterations are therefore a response to changes in the target cells rather than a reflection of zygosity directly impacting phenotype expression. Another study showed that NK receptor expression can depend on the gene copy number of other receptor genes. In particular, the two genes KIR3DL1 and KIR3DS1, coding for killer cell immunoglobulin receptors (KIR), are important in the containment of

HIV-I [34]. Similar to NKG2C, the KIR3DS1 receptor is only expressed by a subset of the NK population. Interestingly, the study by Pelak et al. [34] showed that the fraction of KIR3DS1+ cells and RNA transcription level increased with increasing copies of KIR3DL1. This resembles the observations from model **B**, if interpreted to represent positive feedback via mutual activation of KIR3DS1 and KIR3DL1. A mere positive feedback from KIR3DL1 onto itself and subsequent activation of KIR3DS1 would be represented by model C.

In this study we have only considered one or two gene copies, but some NK receptors belonging to the KIR-family are observed to have three gene copies [35]. The mathematical analysis and the qualitative differences between models A, B and C will still be true for $n_z > 2$. The aim of our models was to understand the differences between donors of different zygosity, but also to provide a rough tool for locating positive feedbacks of bistable phenotypes. Given a bistable phenotype, mathematical modeling can use donor groups of different zygosity to restrict the location of the positive feedback, as we have exemplified with the expression of NKG2C in HCMV positive donors.

Author Contributions

Conceptualization, T.H.H., E.A., A.M., M.L.B., J.G.O.; Methodology, T.H.H., E.A., M.H.J., A.T., J.G.O.; Software, T.H.H., A.T.; Investigation, T.H.H.; Resources, T.H.H., A.M., M.L.B., A.T.; Writing - Original Draft, T.H.H., E.A., J.G.O.; Writing - Review & Editing, T.H.H., E.A., A.M., M.L.B., M.H.J., A.T., J.G.O.; Supervision - M.H.J., A.T., J.G.O.; Funding Acquisition - T.H.H., M.L.B., M.H.J., A.T., J.G.O.

Acknowledgements

This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) and FEDER (projects FIS2015-66503-C3-1-P and SAF2013-49063-C2-1-R), the Center for Models of Life, Niels Bohr Institute, Copenhagen University, and the Red Española de Esclerosis Múltiple (REEM, project RD12/0032/0005 ISCIII-MINECO). J.G.O. also acknowledges support from the ICREA Academia programme, the Generalitat de Catalunya (project 2014SGR0947), and from the "María de Maeztu" Programme for Units of Excellence in R&D (MINECO, project MDM-2014-0370). M.L.-B. and A.M. are also supported by the EU FP7-MINECO Infect-ERA program (project PCIN-2015-191-C02-01).

References

1. Hewitt EW. The MHC class I antigen presentation pathway: strategies for viral immune evasion. Immunology. 2003;110(2):163–169. doi:10.1046/j.1365-2567.2003.01738.x.

2. Lanier LL. NK cell recognition. Annual Review of Immunology. 2005;23(1):225–274. doi:10.1146/annurev.immunol.23.021704.115526.

3. Vivier E, Tomasello E, Baratin M, Walzer T, Ugolini S. Functions of natural killer cells. Nature Immunology. 2008;9(5):503–510. doi:10.1038/ni1582.

4. Sun JC, Beilke JN, Lanier LL. Adaptive immune features of natural killer cells. Nature. 2009;457(7229):557–561. doi:10.1038/nature07665.

5. Guma M, Angulo A, Vilches C, Gómez-Lozano N, Malats N, López-Botet M. Imprint of human cytomegalovirus infection on the NK cell receptor repertoire. Blood. 2004;104(12):3664–3671. doi:10.1182/blood-2004-05-2058.

6. Gumá M, Cabrera C, Erkizia I, Bofill M, Clotet B, Ruiz L, et al. Human Cytomegalovirus Infection Is Associated with Increased Proportions of NK Cells That Express the CD94/NKG2C Receptor in Aviremic HIV-1–Positive Patients. The Journal of Infectious Diseases. 2006;194(1):38–41. doi:10.1086/504719.

7. Lopez-Verges S, Milush JM, Schwartz BS, Pando MJ, Jarjoura J, York VA, et al. Expansion of a unique $CD57^{+}$NKG2Chi natural killer cell subset during acute human cytomegalovirus infection. Proceedings of the National Academy of Sciences. 2011;108(36):14725–14732. doi:10.1073/pnas.1110900108.

8. Costa-Garcia M, Vera A, Moraru M, Vilches C, López-Botet M, Muntasell A. Antibody-Mediated Response of NKG2Cbright NK Cells against Human Cytomegalovirus. The Journal of Immunology. 2015;194(6):2715–2724. doi:10.4049/jimmunol.1402281.

9. Muntasell A, López-Montañés M, Vera A, Heredia G, Romo N, Peñafiel J, et al. NKG2C zygosity influences CD94/NKG2C receptor function and the NK-cell compartment redistribution in response to human cytomegalovirus. European Journal of Immunology. 2013;43(12):3268–3278. doi:10.1002/eji.201343773.

10. Plante E, Mestl T, Omholt SW. Feedback loops, stability and multistationarity in dynamical systems. J Biol Syst. 1995;03(02):409–413. doi:10.1142/s0218339095000381.

11. Becskei A, Séraphin B, Serrano L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. The EMBO journal. 2001;20(10):2528–2535.

12. Bagowski CP, Ferrell JE. Bistability in the JNK cascade. Current Biology. 2001;11(15):1176–1182.

13. Ingolia NT, Murray AW. Positive-feedback loops as a flexible biological module. Current biology. 2007;17(8):668–677.

14. McKinney C, Merriman ME, Chapman PT, Gow PJ, Harrison AA, Highton J, et al. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. Annals of the Rheumatic Diseases. 2007;67(3):409–413. doi:10.1136/ard.2007.075028.

15. Brasch-Andersen C, Christiansen L, Tan Q, Haagerup A, Vestbo J, Kruse TA. Possible gene dosage effect of glutathione-S-transferases on atopic asthma: Using real-time PCR for quantification of GSTM1 and GSTT1 gene copy numbers. Human Mutation. 2004;24(3):208–214. doi:10.1002/humu.20074.

16. Gonzalez E. The Influence of CCL3L1 Gene-Containing Segmental Duplications on HIV-1/AIDS Susceptibility. Science. 2005;307(5714):1434–1440. doi:10.1126/science.1101160.

17. Louis M, Holm L, Sánchez L, Kaufman M. A Theoretical Model for the Regulation of Sex-lethal, a Gene That Controls Sex Determination and Dosage Compensation in Drosophila melanogaster. Genetics. 2003;165(3):1355–1384.

18. Arron JR, Winslow MM, Polleri A, Chang CP, Wu H, Gao X, et al. NFAT dysregulation by increased dosage of DSCR1 and DYRK1A on chromosome 21. Nature. 2006;441(7093):595–600.

19. Espinar L, Dies M, Çağatay T, Süel GM, Garcia-Ojalvo J. Circuit-level input integration in bacterial gene regulation. Proceedings of the National Academy of Sciences. 2013;110(17):7091–7096.

20. Hurvich CM, Tsai CL. Regression and time series model selection in small samples. Biometrika. 1989;76(2):297–307. doi:10.1093/biomet/76.2.297.

21. Wagenmakers EJ, Farrell S. AIC model selection using Akaike weights. Psychonomic Bulletin & Review. 2004;11(1):192–196. doi:10.3758/bf03206482.

22. Brown M, Wittwer C. Flow Cytometry: Principles and Clinical Applications in Hematology. Clinical Chemistry. 2000;46(8):1221–1229.

23. Sun JC, Beilke JN, Bezman NA, Lanier LL. Homeostatic proliferation generates long-lived natural killer cells that respond against viral infection. The Journal of Experimental Medicine. 2011;208(2):357–368. doi:10.1084/jem.20100479.

24. Béziat V, Liu LL, Malmberg JA, Ivarsson MA, Sohlberg E, Björklund AT, et al. NK cell responses to cytomegalovirus infection lead to stable imprints in the human KIR repertoire and involve activating KIRs. Blood. 2013;121(14):2678–2688. doi:10.1182/blood-2012-10-459545.

25. Shu CC, Chatterjee A, Dunny G, Hu WS, Ramkrishna D. Bistability versus Bimodal Distributions in Gene Regulatory Processes from Population Balance. PLoS Computational Biology. 2011;7(8):e1002140. doi:10.1371/journal.pcbi.1002140.

26. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. Nature. 2006;440(7082):358–362.

27. Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene expression. Human Molecular Genetics. 2009;18(R1):R1–R8. doi:10.1093/hmg/ddp011.

28. Ionita-Laza I, Rogers AJ, Lange C, Raby BA, Lee C. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics. 2009;93(1):22–26. doi:10.1016/j.ygeno.2008.08.012.

29. Chang C, Rodríguez A, Carretero M, López-Botet M, Phillips JH, Lanier LL. Molecular characterization of human CD94: A type II membrane glycoprotein related to the C-type lectin superfamily. European Journal of Immunology. 1995;25(9):2433–2437. doi:10.1002/eji.1830250904.

30. Arlettaz L, Villard J, de Rham C, Degermann S, Chapuis B, Huard B, et al. Activating CD94:NKG2C and inhibitory CD94:NKG2A receptors are expressed by distinct subsets of committed CD8+ TCR lymphocytes. European Journal of Immunology. 2004;34(12):3456–3464. doi:10.1002/eji.200425210.

31. Kusumi M, Yamashita T, Fujii T, Nagamatsu T, Kozuma S, Taketani Y. Expression patterns of lectin-like natural killer receptors, inhibitory CD94/NKG2A, and activating CD94/NKG2C on decidual CD56bright natural killer cells differ from those on peripheral CD56dim natural killer cells. Journal of Reproductive Immunology. 2006;70(1-2):33–42. doi:10.1016/j.jri.2005.12.008.

32. Zhao YM, French AR. Two-Compartment Model of NK Cell Proliferation: Insights from Population Response to IL-15 Stimulation. The Journal of Immunology. 2012;188(7):2981–2990. doi:10.4049/jimmunol.1102989.

33. Husain Z, Levitan E, Larsen CE, Mirza NM, Younes S, Yunis EJ, et al. HLA-Cw7 Zygosity Affects the Size of a Subset of CD158b+ Natural Killer Cells. Journal of Clinical Immunology. 2002;22(1):28–36. doi:10.1023/a:1014204519468.

34. Pelak K, Need AC, Fellay J, Shianna KV, Feng S, Urban TJ, et al. Copy Number Variation of KIR Genes Influences HIV-1 Control. PLoS Biology. 2011;9(11):e1001208. doi:10.1371/journal.pbio.1001208.

35. Béziat V, Traherne JA, Liu LL, Jayaraman J, Enqvist M, Larsson S, et al. Influence of KIR gene copy number on natural killer cell education. Blood. 2013;121(23):4703–4707. doi:10.1182/blood-2012-10-461442.