STSM Title: Understanding cell reprogramming as a way to design better cancer drugs

COST STSM Reference Number: TD1409-36454

Period: 2017-04-18 to 2017-04-30

COST Action: TD1409

STSM Applicant: Núria Folguera-Blasco, Centre de Recerca Matemàtica, Barcelona (Spain).

STSM Host: Helen Byrne, Wolfson Centre for Mathematical Biology, Oxford (UK).

23rd May 2017

1 Scientific background and motivation for the visit

In my PhD, we are interested in studying cell reprogramming, that is, the process that differentiated cells undergo in order to become pluripotent cells (stem cells like). This process was discovered by Shinya Yamanaka who, in 2006, found out that by inducing the expression of 4 transcription factors (Oct4, Sox2, Klf4 and c-Myc), a fibroblast from a mouse could be converted into a pluripotent cell [1]. Later on, his lab also showed the reprogramming process to be successful when using a human cell.

This scientific achievement has opened a new area of research and it has boosted the progress of regenerative medicine. However, reprogramming process can also explain the creation of cancer stem cells (CSCs), that is, cancer cells that reprogram towards a less differentiated state, from where they proliferate in an even more harmful way. Moreover, the reprogramming process is inefficient (from those cells who start the reprogramming process, only a few end up being pluripotent), but when the presence of oncometabolites (intermediate products of metabolism associated with cancer) is taken into account, its efficiency increases exponentially [2]. Another key element in the process is aging, because it seems that as cells age, reprogramming is more likely to happen.

Clearly, thus, shedding some light on the reprogramming process can give new insights towards the CSCs understanding, which we believe will be highly valuable, because the study of CSCs is a target in oncology research and, nowadays, little is known about its appearance or maintenance.

Regarding the motivation of the visit, during my PhD, a stochastic cellular model with epigenetic regulation has been developed [3,4]. This model shows that some cells are more prone to reprogramming and, more precisely, these particular cells which are more favour to reprogramming share certain common properties. Interestingly, these same properties have also been associated with aging and cancer, suggesting that reprogramming becomes more likely under these conditions [5]. Some of these properties are related to epigenetic marks, such as the acetylation or methylation levels. The importance of this relies on the fact that epigenetic marks are reversible and therefore represent a promising therapeutic target for cancer and the aging process.

Hence, the idea of the STSM is to formulate a stochastic multiscale model to study the effects of aging and cancer in reprogramming. And working with Professor Byrne seems an excellent opportunity, since she is an expert in multiscale modelling [6,7,8].

2 Work undertaken

During the STSM, we found that the stochastic cellular model with epigenetic model formulated previously had interesting outcomes, but it was not complete in the sense that not all the possible cases were being studied. In particular, the model assumed an abundance in the media of acetyl and methyl groups, which not always hold. More precisely, when there are metabolic disfunctions this is no longer true. Therefore, before going into the stochastic multiscale model, we decided that a generalisation of this previous model in which the acetyl and methyl groups are not abundant could give interesting results, which, in turn, could be more easily linked with metabolism impairment (which is something well documented both in cancer and aging) and hence, at last, with cell reprogramming.

The model itself, then, has these two elements (methyl and acetyl groups) as variables and hence, we let them evolve as a function of time. However, we control them by means of a total amount (i.e, an upper bound, which can be thought as a carrying capacity), and this has to be understood as a parameter whose value can change depending on several circumstances as, for instance, the age of the cell. We believe that this new model will cover a larger range of clinical cases and, in particular, the previous model is incorporated into this one when taking the proper limits. Therefore we are generalising the model previously built, adding extra features to it, which can play an important role.

To be more precise, let's try to briefly depict the model which has been formulated. As variables, we are considering the following:

Table 1: Random variables

Variable	Description
X_1	Number of methylation marks (MM)
X_2	Number of acetylation marks (AM)
X_4	Number of unmodified nucleosomes (U)
X_5	Number of methylated nucleosomes (M)
X_6	Number of acetylated nucleosomes (A)
X_7	Number of HDM enzyme molecules (HDM)
X_8	Number of methylated nucleosome-HDM enzyme complexes (C_M)
X_9	Number of HDAC enzyme molecules (HDAC)
X_{10}	Number of acetylated nucleosome-HDAC enzyme complexes (C_A)
X_{11}	Number of HM enzyme molecules (HM)
X_{12}	Number of methylated nucleosome-HM enzyme complexes ($\tilde{C}_M$)
X_{13}	Number of HAC enzyme molecules (HAC)
X_{14}	Number of acetylated nucleosome-HAC enzyme complexes ($\tilde{C}_A$)

Among these variables, there are several conservation laws that must be taken into account, because they will help to reduce the dimension of the problem afterwards. They read as follows:

$$M + U + A + C_M + C_A + \tilde{C_M} + \tilde{C_A} = N$$

$$C_A + HDAC = HDAC_T$$

$$C_M + HDM = HDM_T$$

$$\tilde{C}_A + HAC = HAC_T$$

$$\tilde{C}_M + HM = HM_T$$

$$M + MM = MM_T$$

$$A + AM = AM_T$$

where, N is the total of nucleosomes, $HDAC_T$ is the total number of HDAC enzyme molecules, HDM_T is the total number of HDM enzyme molecules, HAC_T is the total number of HAC enzyme molecules, HM_T is the total number of HM enzyme molecules and MM_T and AA_T are the total number of methylation and acetylation marks, respectively. It is important to notice that in the model we will assume a separation of scales, where the number of nucleosomes is going to be represented by a larger scale than the number of enzymes, complexes and epigenetic marks (acetylation and methylation). This will allow to suppose that these variables (enzymes, complexes and marks) reach its steady state value faster than the nucleosomes.

Once we have described the variables we are considering in the model, let's see which transitions may happen among them, which defines the set of reactions that our model includes:

- HDM-mediated demethylation:

$$M + HDM \leftrightarrows C_M \rightarrow U + HDM + MM$$

- HM-mediated methylation:

$$U + MM + HM \leftrightarrows \tilde{C}_M \rightarrow M + HM$$

- HDAC-mediated deacetylation:

Table 2: Random processes and their transition rates

Transition rate	Event
$W_5(x) = k_5 X_5 X_7$	Formation of M-nucleosome-HDM enzyme complex (unrec.)
$W_6(x) = k_6 X_8$	M-nucleosome-HDM enzyme complex splits (unrec.)
$W_7(x) = k_7 X_8$	Demethylation and HDM enzyme and MM release (unrec.)
$W_8(x) = k_8 X_5 X_6 X_7$	Formation of M-nucleosome-HDM enzyme complex (rec.)
$W_9(x) = k_9 X_6 X_8$	M-nucleosome-HDM enzyme complex splits (rec.)
$W_{10}(x) = k_{10} X_6 X_8$	Demethylation and HDM enzyme and MM release (rec.)
$W_{13}(x) = k_{13} X_6 X_9$	Formation of A-nucleosome-HDAC enzyme complex (unrec.)
$W_{14}(x) = k_{14} X_{10}$	A-nucleosome-HDAC enzyme complex splits (unrec.)
$W_{15}(x) = k_{15} X_{10}$	Deacetylation and HDAC enzyme and AM release (unrec.)
$W_{16}(x) = k_{16} X_6 X_5 X_9$	Formation of A-nucleosome-HDAC enzyme complex (rec.)
$W_{17}(x) = k_{17} X_{10} X_5$	A-nucleosome-HDAC enzyme complex splits (rec.)
$W_{18}(x) = k_{18} X_{10} X_5$	Deacetylation and HDAC enzyme and AM release (rec.)
$W_{21}(x) = k_{21} X_4 X_1 X_{11}$	Formation of U-nucleosome-HM enzyme-MM complex (unrec.)
$W_{22}(x) = k_{22} X_{12}$	U-nucleosome-HM enzyme-MM complex splits (unrec.)
$W_{23}(x) = k_{23} X_{12}$	Methylation and HM enzyme release (unrec.)
$W_{24}(x) = k_{24} X_4 X_1 X_{11} X_5$	Formation of U-nucleosome-HM enzyme-MM complex (rec.)
$W_{25}(x) = k_{25} X_{12} X_5$	U-nucleosome-HM enzyme-MM complex splits (rec.)
$W_{26}(x) = k_{26} X_{12} X_5$	Methylation and HM enzyme release (rec.)
$W_{27}(x) = k_{27} X_4 X_2 X_{13}$	Formation of U-nucleosome-HAC enzyme-AM complex (unrec.)
$W_{28}(x) = k_{28} X_{14}$	U-nucleosome-HAC enzyme-AM complex splits (unrec.)
$W_{29}(x) = k_{29} X_{14}$	Acetylation and HAC enzyme release (unrec.)
$W_{30}(x) = k_{30} X_4 X_2 X_{13} X_6$	Formation of U-nucleosome-HAC enzyme-AM complex (rec.)
$W_{31}(x) = k_{31} X_{14} X_6$	U-nucleosome-HAC enzyme-AM complex splits (rec.)
$W_{32}(x) = k_{32} X_{14} X_6$	Acetylation and HAC enzyme release (rec.)

$$A + HDAC \leftrightarrows C_A \rightarrow U + HDAC + AM$$

- HAC-mediated acetylation:

$$U + AM + HAC \leftrightarrows \tilde{C}_A \rightarrow A + HAC$$

These reactions can be both recruited and unrecruited, which means that the reaction happens independently of the state of the other nucleosomes (unrecruited) or the transition rate depends on the state of the other nucleosomes (recruited). Transition rates associated with this model are described in Table 2.

From these rates, using law of mass action, we can derive a set of dimensional ODEs for its time evolution. This system is quite large (13 variables) and hence, we will exploit the separation of scales. In particular, it can be assumed that $\frac{dX_7}{dt} \approx 0$, $\frac{dX_8}{dt} \approx 0$, $\frac{dX_9}{dt} \approx 0$, $\frac{dX_{10}}{dt} \approx 0$, $\frac{dX_{11}}{dt} \approx 0$, $\frac{dX_{12}}{dt} \approx 0$, $\frac{dX_{13}}{dt} \approx 0$ and $\frac{dX_{14}}{dt} \approx 0$. This fact, allows us to compute the steady state value of these variables and we can use it into the equations for the other variables. Furthermore, writing $X_1 = MM_T - X_5$ and $X_2 = AM_T - X_6$, we obtain the following system of ODEs:

$$\frac{dX_5}{dt} = \frac{-HDM_T(k_5 + k_8 X_6)(k_7 + k_{10} X_6) X_5}{(k_6 + k_7) + (k_5 + k_8 X_6) X_5 + (k_9 + k_{10}) X_6} + \frac{HM_T(k_{23} + k_{26} X_5)(k_{21} + k_{24} X_5)(MM_T - X_5) X_4}{k_{22} + k_{23} + (k_{25} + k_{26}) X_5 + (k_{21} + k_{24} X_5) X_4 (MM_T - X_5)}$$

$$\frac{dX_6}{dt} = \frac{-HDAC_T(k_{15} + k_{18} X_5)(k_{13} + k_{16} X_5) X_6}{(k_{14} + k_{15}) + (k_{13} + k_{16} X_5) X_6 + (k_{17} + k_{18}) X_5} + \frac{HAC_T(k_{29} + k_{32} X_6)(k_{27} + k_{30} X_6)(AM_T - X_6) X_4}{k_{28} + k_{29} + (k_{31} + k_{32}) X_6 + (k_{27} + k_{30} X_6) X_4 (AM_T - X_6)}$$

where X_4 has not been written although having its own ODE. Instead, what we can do to compute its value (depending on X_5, X_6) is to use one of the conservation laws, that is, $X_5 + X_4 + X_6 + X_8 + X_{10} + X_{12} + X_{14} = N$. Since we know the steady state value for X_8, X_{10}, X_{12} and X_{14}, we have that $X_4 = N - X_5 - X_6 - X_8(ss) - X_{10}(ss) - X_{12}(ss) - X_{14}(ss)$, where ss denotes steady state value.

Therefore, after some algebra, we have an ODE system of 2 equations, that can be studied in detail. Of particular interest is finding the biological realistic steady states and which kind of bifurcations appear.

3 Expected outcomes

As a first outcome from the STSM, what we expect is to analyse in detail the model described so as to study its steady states and its bifurcations as the parameter values vary. This can give valuable information on how different metabolic activities can alter the reprogramming process. Nevertheless, once this model would have been extensively understood, the next step would be to apply the same model but instead of considering a single gene, doing it for a pair of genes, one thought to promote pluripotency and the other thought to promote differentiation. We believe that the system of ODEs that we will obtain will be fairly similar but with 4 coupled equations instead of 2. The reason why these equations will not be decoupled relies on the fact that there are some conservation laws linking them, in particular, the finite number of methylation and acetylation marks, as well as the number of enzymes available.

Resulting from this research, we expect to produce a publication where this model will be described and analysed in detail. Furthermore, we aim to continue the collaboration with Professor Byrne so as to apply the model to other biological systems which she knows in detail (such as the intestinal crypt or determined processes that occur during pregnancy in the development of an embryo). This will enhance the applicability of the model and therefore increase its impact in a broader sense.

As future research plans, we also consider formulating a multiscale model, including to the model presented, a model for protein transcription. Another idea would be to formulate a multiscale model using a metabolic model for the methylation. This would allow to build up a more complete (yet complex) model which will offer a larger picture of the incorporation of the metabolism in the reprogramming process.

4 STSM funding

This research could not have been carried out without the financial help of the STSM funding, because it has allowed me to come to Oxford, meet Professor Byrne and have with her, face to face discussions about the current model and its possible extensions. The STSM has made easier, in particular, the understanding between us and hence, it has boosted the start of this research, taking advantage of Professor Byrne's knowledge on the topic. Therefore, as a conclusion, it can be said that the STSM funding has been a key element in the research process.

References

[1] K. Takahashi, S. Yamanaka. Cell. **126**, 4 (2006).
[2] J.A. Menendez, B. Corominas-Faja, E. Cuyas, M.G. García, S. Fernandez-Arroyo, A.F. Fernandez, J. Joven, M.F. Fraga, T. Alarcon, Stem Cell Reports. **6** (2016).
[3] I. B. Dodd, M.A.Micheelsen,K.Sneppen, and G. Thon, Cell **129**, 813 (2007).
[4] C. Lu and C.B. Thomson, Cell Metabolism **16**, 9 (2012).
[5] L. Mosteiro, C. Pantoja, N. Alcazar, R.M. Marión, D. Chondronasiou, et al., Science **354**, 6315 (2016).
[6]R.D. O'Dea, M.R. Nelson, A.J. El-Haj, S.L. Waters and H.M. Byrne. Math Med Biol **32(3)**: 345-66 (2015).
[7]T. Alarcón, H.M. Byrne, P.K. Maini. Multiscale Model Simul **3(2)**: 440–475 (2005).
[8]F. Spill, P. Guerrero, T. Alarcon, P.K. Maini, H.M. Byrne, J. Math. Biol. **70**: 485 (2015).