Fourth semester report

Botond Kalocsai (kalocsaibotond@gmail.com)

Ph. D. Program: **Statistical Physics, Biological Physics and Physics of Quantum systems**
Supervisor: **Tibor Vellai**
Ph. D. Thesis title: **Reverse-engineering of molecular biological processes of ageing**

Contents

1 Introduction

Our research focuses on the ageing of biological systems, ultimately on human ageing. We explore the biochemical processes that lead to ageing. Ageing is the biochemical process that causes a living organism's system to become inoperable, degrade and die over time, even when kept under ideal conditions. Some have even linked the ageing process to the development of cancer.

So far, there are two main categories of ageing theories: genetic and epigenetic. Genetic ageing theories postulate that changes in genetic information drive the ageing process. Epigenetic ageing theories postulate that abnormal gene expression leads to the ageing process.

The most modern theory describes ageing with the proliferation of retrotransposons in the genome. Transposons are segments of DNA that can change their

position on the hereditary material through the mechanisms of the cellular machinery. Retrotransposons are transposons that use the cellular machinery to copy themselves back and forth across the genome, thereby corrupting the information encoded by the genome. The most modern theory postulates that the proliferating retrotransposons that accumulate over time cause the cell to become dysfunctional, and cause the exponential ageing process [1]. Meanwhile the relocation of DNA transposons cause slower ageing degradation via the error of their cutting mechanism.

We investigate and develop this theory via bioinformatics or computational biology.

2 Research work

In the first semester, I took the initiative to inspect the retrotransposon-related protein interaction data of C. elegans at `wormbase.org`. After some exploratory data analysis and compiling the data into a proper protein interaction network, this approach proved to be a dead end. Unfortunately, there is not enough transposon-linked protein data in the database. Hence, my supervisor redirected me to a computational biology approach to decipher the evolutionary stability of the ageing-causing retrotransposons using computer simulations.

In the second semester, we continued the work and built computer simulations. These computer simulations were bottom-up, hierarchical ageing models, where we inferred the ageing of the population from cellular ageing. We built a model of cellular ageing. In this model, we inferred ageing from the retrotransposon proliferation. We represented the genome as the number of copies (redundancies) of the intelligible sequence and the length of the non-intelligible, junk region. It forms a state vector of the genome. These sequences are genes and retrotransposons. These copy numbers, or genetic redundancies, decrease when destructive retrotransposon insertion occurs. The length of the junk region increases with the length of the insertion subject sequence. The copying and insertion are stochastic in nature. We inferred the insertion probabilities from the sequence lengths and the accessible genome length ratio. From the genetic state of the cell, we inferred its functionality. Using the cellular functionality of the organs, we inferred organ functionality. We inferred the organism's functionality and fitness from organ functionality.

I also investigated the Hydra Vulgaris genome program. This species is renowned for not having an ageing process. Interestingly, it contains a considerable amount

of transposons.

We experimented with plenty of versions of the model. In general, for non-pathological cases, we found that more transposon activity results in quicker death.

Although we found one pathological initial condition class where the cell starts with a vast number of transposon copies and a genome length. The probability of genome disabling exponentially vanishes while the retrotransposon number and junk genome length explode. Hence, cells survive. This case is metabolically impossible. Introducing some bounding mechanism to transposon numbers and junk genome length eliminates this issue.

Although we found one pathological initial condition class: initially, the genome contains a vast number of retrotransposon copies. The probability of genome disabling exponentially decreases. The retrotransposon number and junk genome length explode. Hence, cells survive. This case is metabolically impossible. Introducing some bounding mechanism to transposon numbers and junk genome length eliminates this issue.

By the third semester, we realised the missing component from our simulations. We needed to introduce the cancer into our model. Cancer has two crucial properties for us. Firstly, the chromatin matter in the nucleus of the cancer cells unfolds. The transcription factors can reach the so-far hidden part of the genome. The unfolding is observable via basic bright field microscopy using hematoxylin-eosin staining. The so-far cached or hidden retrotransposons suddenly become active. Secondly, in cancer cells of observed tumours, the Piwi-piRNA pathways are active. [2]. Piwi-piRNA is renowned for explicitly silencing retrotransposon activity in germ-line cells. It is inactive in healthy somatic cells.

At the implementation level, cancer cells have two key properties. Firstly, they divide (clone) into their containing organ with a given probability. Secondly, they negatively affect an organ's or an organism's functionality and fitness. The genesis of cancer cells occurs via the spawning (cloning) of a chosen somatic cell with a given probability. Somatic cells have a cache of inactive transposons. The model adds the content of the cache to the active transposon stock of the cancer cell. Cancer genesis can occur via the placement of cancer cells into the organ of the cancer-spawning somatic cell or a dedicated tumour organ. In the former case, the cancer cell degrades its containing organ. In the latter case, the cancer cell increases the tumour's functionality. However, the tumour's functionality negatively affects the organism. For simplicity, we worked with the latter.

By the fourth semester, it became clear that having cancerisation in our computer model led us to an optimisation problem. Too much retrotransposon activity causes the organisms to die because of ageing. Too low retrotransposon activity

causes the organisms to die because of cancer. It led us to the conclusion that the retrotransposons and their ageing process are an antagonistic agent of cancer! It has considerable implications. It predicts a universal cancer medicine: to kill cancer, it is enough to disable the Piwi-piRNA pathways in the cancer cells.

In parallel, we also commenced developing analytical models. The problem with the numeric computer simulation is that solving optimisation problems is difficult when the function we sample has no analytic formula and takes a long time to evaluate. Plain grid search or random search is not effective. Still if we want to get a good enough approximation, we can apply Bayesian optimisation using Gaussian process regression [3].

We can formulate first-order ordinary differential equations of the genetic redundancies that describe temporal evolution. However, these equations are deterministic and describe the process on average at best. Exploiting the fact that a living cell is a converted quantity of the space of the genetic state vectors, we can formulate a continuity equation for the finding probabilities on this space, analogously to Liouville's equation. A vector field governs the probability density function. Our first-order ordinary differential equations form this vector field. This sophisticated approach embeds the first-order ordinary differential equations into a stochastic model.

On the other hand, the difficulty of precise measurements of the internals of biological systems and the asymptotic deterministic behaviour of ageing, i.e. assured senescence and death, preempt the need for analytical models of such sophistication. If we do not have many observables, then we have to make do with simple phenomenological models of the few observables we have that only aim to describe the asymptotic deterministic behaviour of assured senescence and death. If we presume the beneficial role of retrotransposons as antagonistic agents against cancer proliferation, we may model the asymptotic behaviour using the somatic, cancer cell division potential and retrotransposon proliferation rates.

Lastly, we have a code base for our simulations, which I want to wrap into a Python package and publish on GitHub.

3 Studies in this semester

This semester was study-heavy for me because I had to complete my credit requirements and prepare myself for the complex examination.

To meet my credit requirements, I completed the "Gene silencing, RNA interference" (BIO/05/09E) course. I took this course because RNA interference,

especially the Piwi-piRNA pathways, is crucial for our research.

However, studying for the theoretical part of the complex exam took most of my time and effort. Unfortunately, none of the available exam subjects are explicitly required for my research so far. So studying my chosen subjects was a disjoint endeavour from my research. I picked Biophysics, Bioinformatics and Data-intensive and machine learning methods subjects. Biophysics was the major subject. I made a note and worked out as many of the topics as my time allowed.

References

[1] Nature Neuroscience volume 16, pages529–531 (2013)
Wanhe Li, Lisa Prazak, Nabanita Chatterjee, Servan Grüninger, Lisa Krug, Delphine Theodorou & Josh Dubnau:
Activation of transposable elements during aging and neuronal decline in Drosophila
https://www.nature.com/articles/nn.3368

[2] Nature volume 505, pages353–359 (2014)
Robert J Ross, Molly M Weiner, Haifan Lin: PIWI proteins and PIWI-interacting RNAs in the soma
https://www.nature.com/articles/nature12987

[3] Daniel Lizotte, Tao Wang, Michael Bowling, Dale Schuurmans: Automatic Gait Optimization with Gaussian Process Regression
https://www.ijcai.org/Proceedings/07/Papers/152.pdf