Université Libre de Bruxelles

Analysis of functional and comparative genomics data

BIOL-F-423

Project : Explore mDNA age in human cancers

Author:
Charlotte Nachtegael

Supervisor:
Vincent Detours

June 8 2016

1 DNA methylation age and human cancers

Biological age could explain the differences observed between two people of the same chronological age, but looking much more older or younger. DNA methylation (mDNA) age pretends to be a mean to obtain our biological age by measuring the effect of epigenetic maintenance cumulated with time (Horvath, 2013). One of the questions asked is "does aging protect against cancer ?" We wanted to know if the mDNA is a good measure of biological age, then if we could find a link between the biological age of cancers tissues and clinical factors and finally which pathways were significant in cancers with the biological age as covariable.

For this project, we received a sample of mDNA age of both cancer and healthy tissues from patients suffering of breast invasive carcinoma (BRCA).

2 Material and programs

- R version 3.3.0
- library survival
- GSEA
- libray samr

3 Chronological age vs mDNA age

First, we wanted to infer exactly how well mDNA age can approximate the chronological age. The correlations were calculated with the spearman method.

We began to calculate the correlation of the real age of the patients with the mDNA age computed for their healthy tissues. A good correlation of 0.8678519 was found between the two variables ($p < 2.2e\text{-}16$). This was confirmed by the plot of the real age in function of the mDNA age of the healthy tissues (Fig 1a), where we can see that most of the points are situated on the diagonal of the plot, denoting a linear correlation between the variables.

For the relation between the mDNA age of tumoral tissues compared to the chronological age, nothing significant could be found with a correlation of 0.3536603

(p = 2.008e-11). The plot of the two variables (Fig 1b) confirmed the result, as we could only see a cloud of points with no shape denoting a relation between the variables.

(a) For healthy tissues (b) For tumoral tissues

Figure 1: Plot of the correlation between the mDNA age and the chronological age

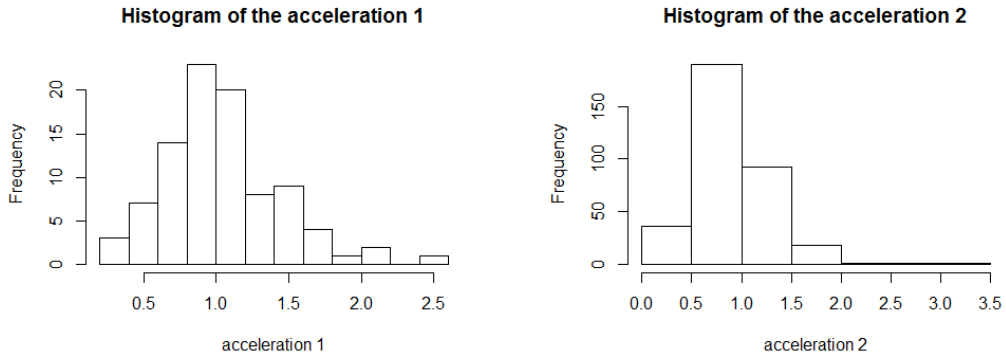
A negative control was performed by permuting the data 100 times and redoing the correlation calculations,further confirming the results were not obtained by hazard.

These results followed the results found in the article of Horvath (2013) where he profess being able to estimate the chronological age with an error of $\pm$ 6.3 years with his mDNA age. Moreover, it was also discussed that the difference between the mDNA age of tumoral tissues and the chronological age was greater than the one with the healthy tissues. This difference, also observed here, could be a mechanism of defence, in the form of the ageing of the cells, against the tumor.

To determine in which capacity the cells interfere with their biological age, we calculated this difference between their normal age and the tumoral age, deemed as acceleration, with two methods :

- $\text{acceleration 1} = \dfrac{\text{mDNA age of tumors}}{\text{mDNA age of healthy tissues}}$
- $\text{acceleration 2} = \dfrac{\text{mDNA age of tumors}}{\text{chronological age}}$

We observed the distribution of both results (Fig 2). We could see that with the first method we obtained a more Normal distribution than with the second. We chose accordingly the first acceleration results to pursue our analysis.



(a) First method (b) Second method

Figure 2: Distribution of the results obtained for the acceleration with both methods

We could observe that in general we witnessed a deceleration instead of an acceleration for the tumoral tissues, which is also observed in the article of Horvath (2013).

4 Does the biological age influence the clinical factors ?

4.1 The biological age

We studied the correlation between the biological age and the clinical factors. We used linear regression to study the categorial variables : the T-, M- and N-stage, the pathological stage, if the patient underwent a radiation therapy and the histological type of the tissue. For the number of invaded lymph nodes, we used a spearman correlation.

Variable	**Correlation/R-squared**	**P-value**
Pathological stage	0.001	0.5594
T Stage	0.01075	0.45
M Stage	0.006278	0.3407
N Stage	0.01815	0.1815
Radiation therapy	0.00543	0.1964
Histology type	0.0661	0.001608
Number of lymph nodes	0.01719764	0.7589

Table 1: Table of the correlation of the mDNA age with clinical factors

We observed only one significant correlation with the histology type of the cancer (Table 1). However, this only explain around 7% of the variance of the mDNA age. As we could not conclude any strong link between the biological age and the clinical factors, we studied if the biological age is a factor for survival with the aid of a survival analysis. We obtained only a p-value of 0.4379867, thus we cannot use the biological age as a signature for survival.

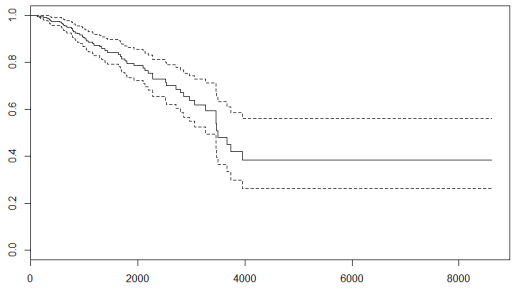


Figure 3: Kaplan-Meier plot for the patients for which we possessed their biological age

4.2 The acceleration

We reiterated the previous calculations, but with the acceleration. Censequently, we had a more restraint sample, as we needed to know the mDNA of both healthy and tumoral tissues to calculate the acceleration (or deceleration) of the ageing of the tumoral tissues (Table 2).

Variable	Correlation/R-squared	P-value
Pathological stage	0.01015	0.8273
T Stage	0.01692	0.6799
M Stage	0.0213	0.3837
N Stage	0.01985	0.779
Radiation therapy	0.01115	0.4026
Histology type	0.04851	0.3576
Number of lymph nodes	-0.01912597	0.8604

Table 2: Table of the correlation of the acceleration with clinical factors

No significant correlation was found between the acceleration values and the clinical factors. For the survival analysis, the p-value for the acceleration was of 0.9092804, also excluding the acceleration as a factor for survival.

5 Which genes and pathways are correlated with the mDNA age and the acceleration in cancer ?

5.1 GSEA

It is a computational method that allows to highlight the significant pathways compared between phenotypes (Subramanian et al., 2005). In this case, we used the mDNA age and the acceleration as a continuous phenotype. The expression matrices chosen were the matrices of normalized expression of the patients for which we had the data for the phenotype (426 for the mDNA age and 92 for the

acceleration). The gene id was used as rows for the expression matrix and the gene name was discarded. We used the following parameters for the run of the GSEA:

- collapse: false
- number of permutations: 1000
- gene set database : c2.cp.v5.1.entrez.gmt
- metric for ranking gene : Pearson

All the results can be found on Dropbox : `https://www.dropbox.com/sh/ubjwmzd6drcb6ak/AAAO1sFerA4C2XDRWlGUSzfPa?dl=0`

5.1.1 mDNA

The top 10 pathways positively correlated with the biological age were primarily involved in metabolism such as fatty acid oxidation, lipid metabolism through the peroxisome and glycan degradation; in membrane structure such as the formation of glycosylphosphatidylinositol and sphingolipids ; and, more interestingly, in the PI3K/mTOR/APK pathway. The latter is deeply connected to the cell cycle and is found activated in cancer, resulting in reduced apoptosis and increased proliferation. So the current hypothesis is that the biological age is linked to an increase of cell metabolism and proliferation, both found in cancer.

The top 10 pathways negatively correlated with the biological age were involved in the inner structure of the cell such as the tubulin and actin pathways, Lisp1 pathway, apparently involved in the nuclear transport, and,more surprisingly, a series of pathways commonly found in other cancers or diseases such as the myc pathway (transcription factor considered as a proto-oncogene), non-small lung cancer, Escherichia Coli infection and glioma pathways. We could extrapolate that biological age protects from some cancer, but perturbs the cytoskeleton in the process.

5.1.2 Acceleration

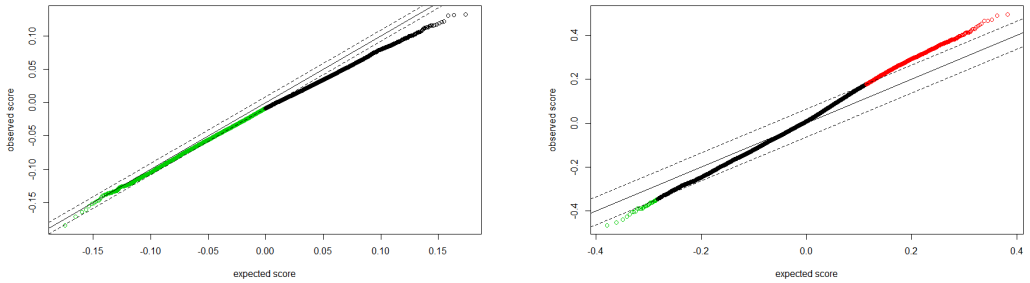
To study more particularly the effect of an acceleration or deceleration of the ageing in tumoral tissues, we did also run GSEA with the acceleration values as phenotype.

The top 10 pathways positively correlated with the acceleration were involved in the structure of the chromatin such as the CTCF pathway; the signalling pathways such as the PAC1 (granule cell survival), SMAD2, alpha S (cAMP-dependent pathway) and trafficking pathways ; the DNA repair with the ATM pathway. This could mean that a high acceleration allows a better protection of the tumoral tissues by increasing their biological age compared to the normal ones.

The top 10 pathways negatively correlated with the acceleration were involved in cell metabolism such as the Krebs cycle, the pyruvate metabolism, the biosynthesis of cholesterol and the formation of pentose phosphate ; and in signalling pathways of immunity cells with the HDAC of class II and the downstream signalling path of B cell receptor. So acceleration would also repress some immunity response and metabolism of cells. The later could act as a deterrent against tumoral cells to limit their proliferation and promote their necrosis.

5.2 SAM

SAM is a method that allows to extract the significant genes expressed in microarray experiments according to specific phenotypes (Tusher et al., 2001); here the biological age and the acceleration. It uses permutations of the data to calculate the rate of false discovery, allowing us to pick not only significant, but also less hazardous results. We used the raw expression matrices of the patients for which we had the phenotype data. The matrices had to be rounded to put put in the *SAMseq* function. All the results can be found on Dropbox : `https://www.dropbox.com/sh/li5l0m1wgbzc2sw/AAAZwWbhN0gHpixT_5vfKCNFa?dl=0`



(a) mDNA (b) Acceleration

Figure 4: Plots of the SAM results

5.2.1 mDNA

No up regulated genes were considered significant enough.

The top 10 down regulated genes coded for : chemokin, protein in regulation of cell development and growth, muscle extra cellular matrix, regulation of cellular differentiation and organogenesis during vertebrate development, promoting TNF-alpha-mediated proteasomal degradation, sex differentiation, accessory subunit of the receptor for interleukin 18 (IL18), a pro-inflammatory cytokin, peptidase, G-protein receptor.

A lot of these genes are involved in the development and the embryogenesis, or inhibits inflammatory behaviour. No significant conclusion can be found according to these results.

5.2.2 Acceleration

We obtained a lot of up regulated genes with the acceleration. The top 10 genes up regulated coded for : protein involved in biological processes involving cell-cell and cell-matrix interactions, including fertilization, muscle development, and neurogenesis ; integral membrane protein in nuclear membrane, component of the survival of motor neurons (SMN) complex, enzyme in sugar metabolism, protein involved in the formation of vesicles from the cellular membrane, regulator of the circadian clock and differentiation of embryonic stem cells, magnesium transporter, transcription factor with no known target, protein involved in vesicle trafficking and a zinc metallopeptidase.

A lot of these genes are involved in the vesicles formation and trafficking and detoxification. No real link with cancer could be found. But this could be because of the great number of results.

The top 10 down regulated genes coded for : an endoplasmic reticulum membrane oxidoreductase, a protein biding to TGF-β, a component of the dystrophin-associated protein complex in the muscle, a protein involved more in the oxidation of fatty acids, a spectrin (cytoskeleton), a ligase promoting the ubiquitination and degradation of the CDK inhibitor (protein involved in the control of the cell cycle), an oxidoreductase required for tumor suppression, a protein involved in the initiation of hepatocyte growth, a protein involved in cytoskeletal reorganization, a protein part of a complex of homologous recombination DNA double-strand break

repair.

Again, we found genes jnvolved in cytoskeleton struture, embryogenesis and development. More interestingly, the gene *HTATIP2*, which is associated with tumor repression, and a protein promoting the degradation of CDKI are down regulated.

6 What are our answers ?

We could not define if there was a link between ageing and protection against tumors as we could not find any correlation with clinical factors or interesting genes and pathways linked to the phenotype.

References

Steve Horvath. Dna methylation age of human tissues and cell types. *Genome Biology*, 14(10):1–20, 2013. ISSN 1474-760X. doi: 10.1186/gb-2013-14-10-r115. URL `http://dx.doi.org/10.1186/gb-2013-14-10-r115`.

Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005. doi: 10.1073/pnas.0506580102. URL `http://www.pnas.org/content/102/43/15545.abstract`.

V. Tusher, R. Tibshirani, and C Chu. Significance analysis of microarrays applied to ionizing radiation response. *Proceedings of the National Academy of Sciences*, 98:5116–5121, 2001.