

Analysing delays between time course gene expression data and biomarkers

Jasmin STRAUBE^{a,b}, Anne BERNARD^a, Bevan Emma HUANG^b and Kim-Anh LÊ CAO^c

^a*QFAB Bioinformatics, Institute for Molecular Biosciences, University of Queensland, QLD 4072, Australia,*

^b*CSIRO Digital Productivity Flagship, Dutton Park, QLD 4102, Australia*

^c*University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, Woolloongabba, QLD 4102, Australia*

Abstract. Associating time course gene expression data to biomarkers can help to understand disease progression or response to therapy. However, detecting associations between these expression profiles is not a trivial task. Often expression changes occur not simultaneously but delayed in time and common used methods to detect correlation will fail to identify these associations. We have developed an efficient approach, DynOmics, based on Fast Fourier Transform to identify coordinated response dynamics between time course ‘omics’ experiments and specific biomarkers of interest while taking time shift into account. We applied DynOmics to a rat study investigating molecular response dynamics to different dosages of acetaminophen (‘paracetamol’). We show how DynOmics can extract relevant molecule expression profiles that enables a better understanding of the molecular pathways related to acetaminophen toxic dosage and renal damage.

Keywords. Time course, gene expression, clinical data, Fast Fourier Transform

Introduction

A biological system adapts to genetic and environmental changes by dynamic interactions and expression changes of its components (e.g. transcripts, proteins, metabolites). The components involved in this adaptation form characteristic profiles over time and similar profiles indicate interactions of components. The challenge is to understand how these interactions give rise to the function and behaviour of that system. One way to reveal molecular responses associated to biological markers is to associate gene expression profiles to known measured biomarkers. The advantage is that clusters of molecule response profiles and potential biological function are predefined rather than unsupervised. The fundamental assumption is that molecules that show similar response profiles are co-regulated, they interact, or they are involved in the same biological processes or molecular functions. The challenge is to identify associations between profiles. Associated molecule expression profiles often do not overlap directly [1], [2]. This can be the result of differences in the kinetics of expression

between molecules that lead to profiles that are delayed in time (Figure 1).

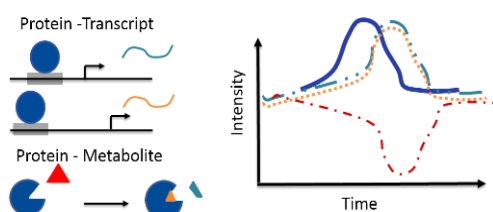


Figure 1. Representation of potential molecule interactions and their quantitative response.

Proteins acting as transcription factors can regulate the expression of multiple genes. Delays occur between the presences of a protein to the transcription initiation of a downstream target. Proteins functioning as enzymes will metabolise substrates resulting in a reduction of that substrate while the quantities of the products increase.

Several authors observed time delays between their expression profiles when integrating transcript and metabolite data [3], [4]. By incorporating time shifts, they could improve on the quality of biological inference. Takahashi *et al.* [3] studied transcript metabolite associations using lagged Pearson correlation. Lagged Pearson correlation maximises the correlation between two profiles while shifting them along time. However, it does not seem to be very sensitive to a small number of time points. Similar issues were observed using Dynamic Time Warping (DTW) [4] to identify multi ‘omics’ associations.

We therefore developed an algorithm based on the Fast Fourier Transform (FFT). The advantage of our approach is that it is still very sensitive for low numbers of time points, a commonly encountered characteristic in time course experiments. It also enables the estimation of time delay between molecules. This information can be used to identify and compare clusters of delayed molecules to find differences in response dynamics.

Possible applications of DynOmics include the comparison of patients’ response to different types of treatment, or monitoring disease progression by associating biomarkers to gene expression data in time course experiments. In this study, we apply DynOmics to acetaminophen toxicity time course data from rats. Our analysis shows how acetaminophen dosage has an impact on metabolism and enables a better understanding of acetaminophen toxicity related to liver and renal damage.

1. Material and methods

1.1. DynOmics algorithm

We developed DynOmics, an algorithm to associate molecule expression profiles taking delay into account, using FFT. FFT decomposes profiles into circular components. This circle information like the frequency (speed), the amplitude (size) and the phase angle (delay), can be used to compare profiles. To date, people have used information about the frequency to cluster cell cycle genes [5]. We developed a method to use the phase angle of the dominant frequency of the profiles to estimate the delay between profiles. This delay was then used to realign profiles and Pearson’s correlation was used to assess their relevance. Profiles identified as associated were then further analysed via gene enrichment analysis.

1.2. Acetaminophen toxicity data

Acetaminophen (‘paracetamol’) is one of the most used pain relievers and fever reducers. However, acetaminophen overdose has severe consequences like liver and renal damage and liver failure. Here we examined a study performed by Bushel *et al.* [6] to analyse the impact of acetaminophen dosage in association with biomarkers and gene expression in rats. Microarray gene expression data of 3316 genes were obtained from four male Fischer rats per dose group exposed to subtoxic levels (50 and 150 mg/kg) and toxic levels (1500 and 2000 mg/kg) of acetaminophen. Rats were sacrificed

to obtain the liver samples at time points 6, 18, 24, or 48 hours after treatment. In addition for each rat clinical chemistry measurements were taken as estimate of acetaminophen damage. Amongst others a biomarker for renal damage, blood urea nitrogen (BUN) was measured. Our analysis focused on the molecular response dynamics to different doses of acetaminophen in association with the BUN biomarker. In particular, we asked whether different delays in gene expression can be attributed to different acetaminophen dosages and whether those delayed genes are involved in relevant molecular pathways with respect to renal damage and acetaminophen metabolism.

1.3. Data pre-processing, modelling and analysis

Data were pre-processed as described in [6]. The low number of time points and high noise hampered the identification of associations of biomarker to the gene expression data. Therefore, we used the spline modelling from the R package *lmms* [7] to summarise the profiles for each dose group separately and interpolated 14 time points regularly spaced between 6 and 48 hours. *lmms* models profiles depending on their variance structure ranging from simple linear models to spline models taking subject-specific intercept and slope into account. Profiles that could only be modelled by a linear model were removed prior to analysis leaving 401 genes for 50, 910 for 150, 644 for 1500, and 960 for 2000 mg/kg for subsequent DynOmics analysis. Genes were termed ‘associated’ if the absolute Pearson correlation of the realigned profiles was above 0.9. Enrichment analysis of associated genes was performed using QIAGEN’s Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity).

2. Results

2.1. DynOmics in comparison to other methods on simulated data

We compared DynOmics to various proposed methods to associate profiles (DTW4Omics [4], Pearson and lagged Pearson correlation) on simulated data sets with either 7 or 14 time points, different noise levels ($\sim N(0, \sigma^2)$; $\sigma = 0.1, 0.2, 0.3, 0.5$) and different delays in the profiles. For all methods we called profiles associated when the Pearson correlation coefficient was above 0.9. Only DTW4Omics provides a p-value from a permutation test for associations of aligned profiles and an association was called when the FDR adjusted p-value was less than 0.05. For 14 time points and low noise levels ($\sigma = 0.1, 0.2, 0.3$), all methods had very high sensitivity and specificity (>0.94) with the exception of Pearson’s correlation (sensitivity <0.5). For 7 time points we observed more differences between the methods. Best performance was observed for DynOmics with sensitivities decreased when the level of noise increased (0.98, 0.94, 0.88, 0.65 respectively). Lagged Pearson was at least 9% less sensitive in all noise scenarios compared to DynOmics (0.89, 0.79, 0.7, 0.51). DTW4Omics (0.6, 0.56, 0.5, 0.26) and Pearson (0.35, 0.32, 0.29, 0.16) showed poor sensitivity. Overall, specificity was stable and high (>0.9) for all methods when noise level increased. To summarise, DynOmics showed high sensitivity and specificity for a low number of time points compared to other methods, an important advantage as time course experiments usually include a small number of time points.

2.2. Expression dynamics in response to renal damage induced by acetaminophen

Having established that DynOmics was competitive to other existing approaches, we next investigated molecular response to different dosages of acetaminophen in the acetaminophen toxicity study. The aim was to detect associations between the BUN biomarker for renal damage and time course gene expression data. We observed that BUN quantity changed temporarily over time for all dosages (Figure 2), with differences in the rate of change. The rate of change increased with increasing acetaminophen dosage, indicating severe renal damage with higher dosage and faster molecular responses. This temporary response indicates only transient renal dysfunction or damage. Due to the temporary change of the BUN marker we expected to identify associated genes that are involved in renal damage but also in the metabolism of acetaminophen.

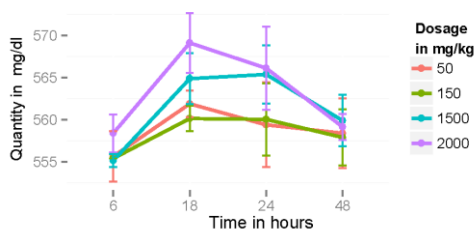


Figure 2. Renal damage marker expression profiles. Presented is the quantity of the BUN renal damage marker in dependency of time after varying dosages of acetaminophen. Points indicate the mean response of the biological replicated measurements and error bars the 95% confidence interval.

Application of the DynOmics algorithm to the level of BUN of varying dosage of acetaminophen resulted in different numbers of associated genes to the renal damage marker BUN. Using a correlation threshold of 0.9, we obtained 68 (22) positive (negative) correlated profiles for 50, 495 (80) for 150, 440 (75) for 1500 and 345 (118) for 2000 mg/kg acetaminophen. Here positive associations indicated the same trend as BUN in Figure 2, therefore these genes were up-regulated. Accordingly, negative associated genes to BUN indicated a down-regulation. Histograms of the predicted delays separated by dosage and correlation indicated that some delays occurred more often than others (Figure 3). We also observed that these prevalent delays are different for subtoxic (50 and 150 mg/kg) and toxic dosage (1500 and 2000 mg/kg) of acetaminophen. Moreover, patterns varied between positive (Figure 3A) and negative (Figure 3B) correlated profiles, suggesting different molecular response dynamics and interactions, with potential different biological implications.

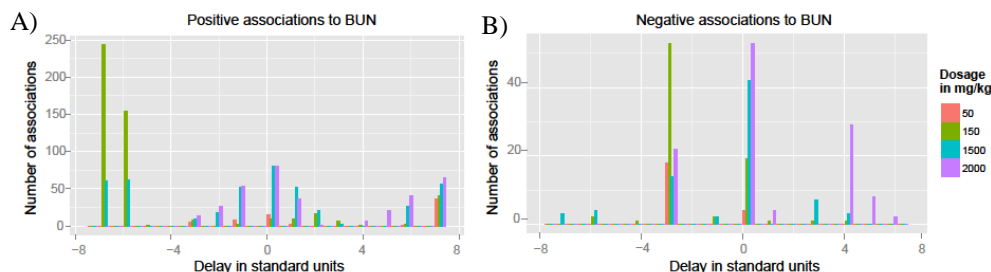


Figure 3. Histograms of predicted delays of the associated gene expression profiles A) positive correlated profiles and B) negative correlated profiles to the renal damage marker BUN. The x-axis presents the delay in standard units where 1 unit refers to 3.2 hours. Negative (positive) delay indicates the BUN marker changed expression prior to (after) the associated gene.

2.3. Biological relevance of shifts and correlation

The biological implication of delayed responses is important to understand signalling pathways and malfunctions. Information obtained can be used to understand the molecular processes involved in renal damage caused by acetaminophen or the detection of potential renal damage prior to the actual event. Therefore, we further investigated the significant enriched pathways (p-value <0.05) or functions from the toxicity list using IPA (Table 1). Associated genes for 2000 and 1500 mg/kg were separated into positive or negative associations to the renal damage marker BUN. We also investigated whether changes in BUN quantity occurred before (negative delay), after (positive delay) or simultaneously (no delay) to gene expression changes.

Table 1. Significant enriched toxicology functions and pathways of associated genes to the BUN renal damage marker identified by DynOmics for at least one dose group. The table indicates the number of genes associated their direction and delay. '+ corr' and '- corr' indicate the number of genes positively or negatively correlated to BUN. A negative (positive) delay indicates that the BUN marker changes prior to (after) the change in gene expression, 'No Delay' indicates that BUN and associated gene changes simultaneously. Dosages without associated genes to the according IPA Tox List were not listed.

Dosage in mg/kg	IPA Tox List	# of genes	+ corr	- corr	- delay	No delay	+ delay
2000	Xenobiotic Metabolism	14	9	5	1	6	7
1500	Signalling	18	13	5	7	7	4
150		20	17	3	16	1	3
50		1	1	-	-	-	1
2000	NRF2-mediated	14	12	2	2	7	5
1500	Oxidative Stress	17	13	4	4	10	3
150	Response	19	17	2	16	-	3
50		4	3	1	1	-	3
2000	Renal Necrosis/Cell	16	15	1	8	4	4
1500	Death	20	18	2	5	9	6
150		19	16	3	17	1	1
50		2	1	1	1	-	1
2000	Long-term Renal Injury	3	3	-	2	-	1
1500	Anti-oxidative Response	3	3	-	-	3	-
150	Panel (Rat)	2	2	-	2	-	-
2000	Increases Renal Damage	4	3	1	1	2	1
1500		2	2	-	1	1	-

The xenobiotic metabolism was clearly activated through dosages of acetaminophen. Differences in the response dynamics were observed for high and low dosages. While molecules with a high dosage reacted prior to or simultaneous to the renal damage marker indicating a fast detoxification of the drug, 150 mg/kg dosage showed late response.

Increased metabolism of acetaminophen can lead to the toxic byproduct *N*-acetyl-*p*-benzoquinone imine (NAPQI) [8]. NRF2 has been shown to play an important role in the regulation of NQO1 which metabolises the hepatotoxic acetaminophen intermediate NAPQI [9]. Majority of molecules involved in the NRF2-mediated oxidative stress response for toxic dosage of acetaminophen were positively correlated with the renal damage marker and occurred simultaneously, indicating the activation of the NRF2 cascade to regulate presence of NAPQI. However, changes for 150 mg/kg occurred later than changes in the BUN renal damage marker which indicated a delayed response due to less presence or slower accumulation of NAPQI. The number of genes identified to be associated to our renal damage marker and known to play a role in

'Renal Necrosis/Cell Death' and 'Long-term Renal Injury Anti-oxidative Response Panel (Rat)' decreased with decreasing dosage, indicating less severe damage for lower dosages. The majority of profiles were positively associated but dynamics varied across the marker. Genes known to increase renal damage only occurred for toxic dosage of acetaminophen.

3. Conclusion and future work

Associated molecular events often do not change simultaneously but are delayed in time. We presented our method DynOmics as a useful and important tool to identify time delayed associations and to understand pathophysiological events.

In this case study we showed gene expression changes caused by acetaminophen and associated to renal damage showed different response dynamics depending on the dosage. Our approach enabled the identification of relevant activated pathways and genes that were delayed in time, giving insights into acetaminophen metabolism. These associations would have been not detected when performing an analysis not taking time shifts into account. This information can be used for early detection of acetaminophen over-dosage and prevention of renal damage.

We are currently investigating the application of DynOmics to detect associations between multiple time course 'omics' data sets (transcriptomics, proteomics, metabolomics, etc.) to understand molecular interactions between multiple molecular functional levels. We will use the information on whether molecules are positively and negatively correlated and on whether they are delayed in time to build and compare directed molecular networks in order to visualise and understand disease at a molecular level. DynOmics is implemented in R and will be made available on CRAN for easy access and use.

References

- [1] W. A. Schmitt, R. M. Raab and G. Stephanopoulos, Elucidation of gene interaction networks through time-lagged correlation analysis of transcriptional data, *Genome Research* **14** (2004), 1654-1663
- [2] J. Qian, M. Dolled-Filhart, J. Lin, H. Yu and M. Gerstein, Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions, *Journal of molecular biology* **314** (2001), 1053-1066
- [3] H. Takahashi, R. Morioka, R. Ito, T. Oshima, M. Altaf-Ul-Amin, N. Ogasawara and S. Kanaya, Dynamics of time-lagged gene-to-metabolite networks of *Escherichia coli* elucidated by integrative omics approach, *Omics: a journal of integrative biology* **15** (2011), 15-23
- [4] R. Cavill, J. Kleinjans and J.-J. Briede, DTW4Omics: Comparing Patterns in Biological Time Series, *PLoS one* **8** (2013), e71823
- [5] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein and B. Futcher, Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization, *Molecular biology of the cell* **9** (1998), 3273-3297
- [6] P. R. Bushel, R. D. Wolfinger and G. Gibson, Simultaneous clustering of gene expression data with clinical chemistry and pathological evaluations reveals phenotypic prototypes, *BMC Systems Biology* **1** (2007), 15
- [7] J. Straube, D. Gorse, B. E. Huang and K.-A. Lê Cao, A linear mixed model spline framework for analyzing time course 'omics' data, (**Submitted**) (2015),
- [8] R. P. Bender, R. H. Lindsey, D. A. Burden and N. Osheroff, N-acetyl-p-benzoquinone imine, the toxic metabolite of acetaminophen, is a topoisomerase II poison, *Biochemistry* **43** (2004), 3731-3739
- [9] L. M. Aleksunes and J. E. Manautou, Emerging role of Nrf2 in protecting against hepatic and gastrointestinal disease, *Toxicologic pathology* **35** (2007), 459-473