

Chapter 32

Chemosensitivity Prediction of Tumours based on Expression, miRNA, and Proteomics Data

I. Tsamardinos

ICS-FORTH, Greece & University of Crete, Greece

G. Borboudakis

ICS-FORTH, Greece & University of Crete, Greece

E. G. Christodoulou

BIOTEC Technische Universität Dresden, Germany

O. D. Røe

Levanger Hospital—Nord-Trøndelag Health Trust, Norway, & Norwegian University of Science and Technology, Norway

ABSTRACT

The chemosensitivity of tumours to specific drugs can be predicted based on molecular quantities, such as gene expressions, miRNA expressions, and protein concentrations. This finding is important for improving drug efficacy and personalizing drug use. In this paper, the authors present an analysis strategy that, compared to prior work, retains more information in the data for analysis and may lead to improved chemosensitivity prediction. The authors apply improved methods for estimating the GI50 value of a drug (an indicator of the response to the drug), regression methods for constructing predictive models of the GI50 value, advanced variable selection techniques, such as MMPC, and a multi-task variable selection technique for identifying a small-size signature that is simultaneously predictive for several drugs and cell lines. The methods are applied on gene expression, miRNA expression, and proteomics data from 53 tumour cell lines after treatment with 120 drugs, obtained from the National Cancer Institute databases. A biological interpretation and discussion of the results is presented for the most clinically important subset of 14 drugs.

DOI: 10.4018/978-1-4666-3604-0.ch032

INTRODUCTION

Prior work shows that the sensitivity of a tumour to a drug can be predicted better than chance based on the gene-expressions of the tumour (Potti et al., 2006; Augustine et al., 2009). This finding paves the way to personalized therapy models. In addition, identifying the molecular quantities that are predictive may lead to a better understanding of the biological mechanisms a drug employs to attack the tumour. In this paper, we develop an analysis strategy to produce predictive models, estimate their performance, and identify the smallest, most-predictive set of molecular quantities required for prediction. The strategy is first applied to and evaluated on the prediction of the response to a set of 120 chemotherapeutic agents based on the responses measured on 53 solid-tumour cell lines; the data have been obtained from the National Cancer Institute databases and contain pre-treatment gene expression, miRNA expression, and protein concentration profiles of the cell-lines. Subsequently, we focus our interest on a subset of 14 drugs that are the most interesting in clinical practice and provide a detailed presentation and biological interpretation of the results. In addition, we apply a method for multi-task feature selection which selects molecular quantities that combined, are simultaneously predictive for an array of drugs. Such algorithms are important for selecting the optimal therapy, by being able to predict the response to several drugs at once by measuring only a small set of molecular quantities.

Compared to prior works, our proposed strategy differs in several ways. The machine learning and statistical analysis employed in the literature process the data in a way that reduces the available information with potential detrimental effects both on the models' prediction performance as well as the identification of the molecular signatures (Potti et al., 2006; Augustine et al., 2009; Staunton et al., 2001). First the estimation of the response to a drug in prior work maybe sub-optimal (Potti et al., 2006; Augustine et al., 2009). The response of

a tumour depends of course, on the dosage. The National Cancer Institute has treated a panel of 60 cancer cell lines with several thousand drugs and has created a dosage-response profile for each combination of drug and tumour. Often, this profile is summarized with a single value such as the $\log_{10}$ GI50. The GI50 stands for "growth inhibition 50%," the concentration of a given test drug that causes 50% growth inhibition at 48 hours, corrected for the cell count at time zero. NCI, in the majority of cases, estimates $\log_{10}$ GI50 by piecewise linear interpolation which are then employed by all prior work (e.g., Potti et al., 2006; Ma et al., 2009; Staunton et al., 2001). In this paper, we show that estimating the $\log_{10}$ GI50 values by fitting a sigmoid to the dosage-response profile preserves more information about the effects of the drug that lead to statistically significantly improved predictive performance.

Second, prior work typically quantizes the $\log_{10}$ GI50 values to create classes of tumours: Potti et al. (2006) and Augustine et al. (2009) categorize tumours as sensitive and resistant, while Staunton et al. (2001) and Ma et al. (2009) as sensitive, intermediate, and resistant. This type of quantization allows the application of machine learning classification techniques, variable selection methods for classification tasks, and statistical hypothesis testing techniques for discrete outcomes. However, our computational experiments demonstrate that maintaining the exact $\log_{10}$ GI50 values and employing regression analysis instead of classification is often preferable as it improves chemosensitivity prediction in approximately half of the cases.

Third, prior work often employs simple methods for identifying molecular signatures such as selecting the top k genes that are mostly differentially expressed between different classes of tumours. We show that more sophisticated methods such as the Max Min Parents and Children (MMPC) algorithm for multivariate feature selection (Tsamardinos, Brown, & Aliferis, 2006)

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/chemosensitivity-prediction-tumours-based-expression/76085

Related Content

The Avatar as a Self-Representation Model for Expressive and Intelligent Driven Visualizations in Immersive Virtual Worlds: A Background to Understand Online Identity Formation, Selfhood, and Virtual Interactions

Colina Demirdjian and Hripsime Demirdjian (2020). *International Journal of Applied Research in Bioinformatics* (pp. 1-9).

www.irma-international.org/article/the-avatar-as-a-self-representation-model-for-expressive-and-intelligent-driven-visualizations-in-immersive-virtual-worlds/261865

Translation of Biomedical Terms by Inferring Rewriting Rules

Vincent Claveau (2009). *Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration* (pp. 106-123).

www.irma-international.org/chapter/translation-biomedical-terms-inferring-rewriting/23057

Natural Knowledge of Smart Bioinformatics to Reduce Tasks Without Added Value or Human Contact in a Pandemic

Potapova Irina (2022). *International Journal of Applied Research in Bioinformatics* (pp. 1-4).

www.irma-international.org/article/natural-knowledge-smart-bioinformatics-reduce/290342

Low Dose Pioglitazone Attenuates Oxidative Damage in Early Alzheimer's Disease by Binding mitoNEET: Transcriptome-To-Reactome™ Biosimulation of Neurons

Charles D. Hammack, George Perry, Richard G. LeBaron, Greg Villareal and Clyde F. Phelix (2015). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 24-45).

www.irma-international.org/article/low-dose-pioglitazone-attenuates-oxidative-damage-in-early-alzheimers-disease-by-binding-mitoneet/165548

Overview of Biomedical Engineering

Ahmad Taher Azar (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications* (pp. 1-28).

www.irma-international.org/chapter/overview-biomedical-engineering/76054