Personalized drug repositioning using gene expression

Koudijs, K.K.M.

Citation

Koudijs, K. K. M. (2023, June 6). *Personalized drug repositioning using gene expression*. Retrieved from https://hdl.handle.net/1887/3619741

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3619741

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 9

Summary

Drug repositioning (also called drug repurposing) is the use of existing drugs for new diseases. Transcriptome Signature Reversion (TSR) is a widely used drug repositioning method to find new anti-cancer drugs. The rationale behind TSR is that the potential of drugs to reverse the gene expression of the disease is predictive of how therapeutically useful it would be in treating that disease. To perform TSR to find new potential anti-cancer drugs, you need three things: 1) a tumor gene expression signature, 2) a drug gene expression signature and 3) a method to combine the two to quantify how much the drug gene expression signature is expected to reverse the tumor gene expression signature.

To create a tumor gene expression signature, typically a differential expression analysis is performed comparing the average gene expression of the tumor tissue samples belonging to specific tumor type (e.g., colon cancer) to the gene expression of adjacent normal tissue samples. Often an additional cutoff is applied to only include genes in the tumor gene expression signature which have the most statistical significance or the difference in expression exceeds a specific threshold (e.g., the gene should be at least twice as much or as less expressed in the tumor tissue samples). To create a drug gene expression signature a differential expression analysis is performed between the gene expression of cell lines after drug exposure and unexposed cell lines. The most common way to calculate the expected reversion of the tumor signature is called the connectivity score method. This method produces a connectivity score for each tumor gene expression signature – drug gene expression signature pair which ranges from -1, i.e., complete reversion of the tumor gene expression signature expected, to +1, i.e., the genes of the tumor gene expression signature and the drug gene expression signature are differentially expressed in the same direction and expected to amplify each other. In summary, TSR predicts that drugs which produce a negative connectivity score with the tumor gene expression signature of a specific tumor type have potential to be therapeutically used against that tumor type.

The thesis starts with **chapter 1**, which is a general introduction to drug repositioning. Here, the need for drug repositioning, its increasing popularity and available methods are described. We discuss how laboratory experiments, clinical observation/retrospective observational studies and prospective studies have been and are being used to find new anti-cancer drugs. Finally, we briefly discuss computational methods which rely on other principles than TSR and introduce TSR at the end.

The TSR methodology and key TSR studies are critically reviewed in **chapter 2**. The databases which have been used to create the drug signatures (CMAP and LINCS L1000) are discussed. In addition, the most cited TSR studies and key conclusions which can be drawn from them are reviewed. In addition, it is discussed why and how drug repositioning could benefit from an individualized approach and what are challenges and potential computational improvements in the application of TSR.

In **chapter 3** the results from the individualized drug repositioning approach based on TSR using 534 clear cell renal cell carcinoma (ccRCC) tumor tissue samples and 72 matched adjacent normal tissue samples are described. Three different methods to create tumor signatures were benchmarked: 1) compare the average ccRCC sample to the 72 adjacent normal control samples (average tumor gene expression signature), 2) compare the average of the 4 ccRCC subtypes to the 72 adjacent normal samples (4 subtype tumor signatures) and 3) compare the individual ccRCC tumor samples to the 72 adjacent normal samples (individual tumor signatures). It is demonstrated that the individual tumor signatures outperform the average and subtype tumor signatures on various metrics, including how many drug signatures show negative enrichment and retrieval of existing targeted drugs such as sirolimus and temsirolimus.

The gene expression data we used in **chapter 3** was generated using bulk RNA-seq, which measures the mRNA transcripts from all cell types present in the sample simultaneously. This complicates interpretation and may have led to inaccurate results. **Chapter 4** expands upon the results in **chapter 3** by performing a sensitivity analysis of the results. The connectivity scores of the top 8 drugs generated using the individual tumor signatures are plotted against the estimated fraction of tumor cells in each sample to diagnose whether the connectivity scores become more negative (i.e., more expected reversion) as the fraction of tumor cells in the tissue samples increases. Surprisingly, the connectivity scores become neutral as the fraction of tumor cells in samples approach 100%. Further analysis revealed that the likely explanation is that the dataset was contaminated with chromophobe and papillary RCC samples, which contain a much higher fraction of tumor cells than the typical ccRCC sample and show much different patterns of negative enrichment.

Chapter 5 describes the systematic validation of TSR for the purpose of anti-cancer drug repositioning with the data of 18 different solid tumor types. Surprisingly, the results show that TSR in the way it has been used does not offer the predictive performance as reported previously in the literature on an earlier validation with 3 solid tumor types. Even worse, it is proven that the reversion of the tumor signature is only a proxy of the general anti-proliferative effect of a drug and therefore does not offer any specificity in prioritizing which drugs may be effective against any specific tumor type. This finding thus invalidates much of the earlier research into TSR and the implications of this are discussed.

One of the assumptions underlying standard RNA-seq normalization is that most genes not differentially expressed across samples and violation of this assumptions reduces statistical power and may result in false positive results. **Chapter 6** describes a new method of normalizing the gene expression of tumor samples using a panel of reference genes which are automatically selected using a variance criterion. We validate this new method by associating the sample correction factors generated using

the new method with biological tumor characteristics and contrasting these with results of the correction factors generated using traditional mRNA-seq normalization methods such as trimmed mean of M values (TMM), trimmed mean of M-values with singleton pairing (TMMwsp) and upperquartile normalization. **Chapter 7** explores whether convolutional neural networks (CNNs) offer any benefits to non-convolutional neural networks when applied to gene expression data.

The thesis ends with a general discussion in **chapter 8**. The emphasis of the discussion is on the current use of TSR as a drug repositioning method, which issues can explain the poor predictive performance and what can be done to make it a more useful drug repositioning method in the future.