OXFORD UNIVERSITY PRESS | Bioinformatics

teff: estimation of Treatment EFFects on transcriptomic data with casual random forest

Journal:	*Bioinformatics*
Manuscript ID	BIOINF-2021-2268
Category:	Applications Note
Date Submitted by the Author:	21-Oct-2021
Complete List of Authors:	Caceres, Alejandro; Instituto de Salud Global Barcelona, Bioinformatics Research Group Gonzalez, Juan; Instituto de Salud Global Barcelona, Bioinformatics Research Group in Epidemiology
Keywords:	

Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Manuscript Category

Subject Section

teff: estimation of Treatment EFFects on transcriptomic data with casual random forest

Alejandro Cáceres* and Juan R. González

Instituto de Salud Global de Barcelona (ISGlobal), Barcelona 08003, Spain

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Causal inference on high dimensional feature data can be used to find a profile of patients who will benefit the most from treatment rather than no treatment. However, there is a need for usable implementations for transcriptomic data. We developed *teff* that applies random causal forest on gene expression data to target individuals with high expected treatment effects.
Results: We applied *teff* to extract a profile of high benefit of treating psoriasis with brodalumab and observed that it was associated with high T cell abundance in non-lesional skin at baseline and a low response to etanercept in an independent study. Individual patient targeting with causal inference profiling can inform the choice between treatments before the intervention begins.
Availability and Implementation: *teff* is an R package available at https://github.com/teff-package/teff.
Contact: alejandro.caceres@isglobal.org
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

An important aim of transcriptomic analyses in clinical trials is to determine a set of gene transcripts with significant differences between treated and not treated patients, or between responders and non-responders to a given drug. The analyses offer important biological information of the disease state and the mechanisms of drug action; see for instance some relevant applications to psoriasis (Correa da Rosa *et al.*, 2017; Tomalin *et al.*, 2020; Zaba *et al.*, 2009; Wang *et al.*, 2020). When evaluating treatment response, differential gene expression analyses are often applied to the treated group only, and not across treatments. As such, the analyses disregard that individuals under placebo can also improve. However, in the interest of helping a patient choose between taking a drug or not, we want to find the genes whose transcription levels provide the highest differences in response between treatment and placebo. While we can perform transcriptome-wide associations for the interaction between treatment and response, one of the main goals in personalized medicine is to characterize patients and assesses their expected individual benefit in following a treatment rather than another one (e.g. placebo) before the intervention begins.

These estimations require the application of causal inference methods that deal with unobserved counterfactual data. Latest machine learning methods in causal inference can be adapted to clinical trials to consider the specificities of transcriptomic data, and then estimate the expected individual treatment effect of each patient; that is, the expected difference in response between treatment and placebo for each individual patient at baseline. However, the methods still need to be implemented in usable software for this type of data, which requires adjustment for covariates and surrogate variables and a targeting strategy. We, therefore, implemented causal random forest (CRF) (Athey *et al.*, 2019) for transcriptomic data and developed a strategy to target individual patients with high expected benefits of treating their disease with a given drug rather than placebo, before intervention begins. We applied the method on a clinical trial of brodalumab on psoriasis patients to profile those with high benefits to the drug, determine their biological correlates, and forecast their actual response levels when treated with brodalumab or with etanercept, a different biological treatment of psoriasis.

2 Implementation

We implemented *teff* that can be used to target individuals with the highest expected difference in response between drug and placebo. Figure 1 shows a schematic representation of the causal inference and targeting. The method has three main stages. In the first stage (Figure 1A), informative transcripts are selected such that their expression levels are significantly associated, at the transcriptome level, with the interaction between treatment (case-control, drug-placebo) and the response (outcome difference between baseline and follow-up). This can be performed with usual packages such as *limma* and *sva*, which corrects for surrogate variables. In the second stage (Figure 1B), *teff* uses causal modeling to select individuals with high expected response differences between treatments given their expression levels across selected transcripts. Causal modeling is performed with CRF that estimates the expected treatment effect for each individual with its confidence interval. This is performed with the function *predicteff* on formatted feature and treatment effect data, which can include surrogate variables and potential confounders. In *teff*, residual

expression data corrected by surrogate variables and confounders, are used as feature data for CRF. Individual treatment effects can be plotted with *plotPredict*. In the third stage (Figure 1C), a consensus binary profile of expression levels is built (argument: *profile*=TRUE) from the individuals with significant treatment effects, and allowing for further feature selection. The profile can be used to target individuals in other studies with transcriptomic data using the function *target* (Figure 1D). Covariates for adjusting the transcription levels in the target study can be included in the treatment effect data and used in the targeting by passing their names in the parameter *nmcov*. A *plot*=TRUE argument in *target* displays the targeted individuals by the profile with predefined match accuracy (e.g. *match*=0.6). The *target* function can also test the significance of the interaction between the targeting and the treatment on the effect. A range of different models can be used in the parameter *model*. See further details in **Supplementary material** and https://rpubs.com/jrgonzalezISGlobal/teff.

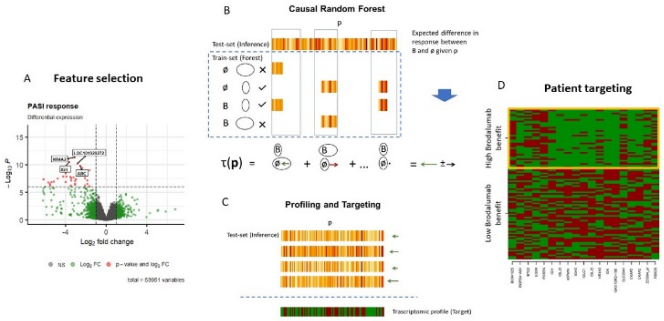

Fig. 1. Steps of *teff* for profiling and targeting. A) Selection of transcripts (red dots) from significant associations with the interaction between response and treatment (drug B/placebo ø). **B**) For each patient in a test-set, estimation of her expected treatment effect τ(p): the expected difference in response (green arrow) between B and ø given the patient transcript levels across the selected transcripts (**p**). τ(p) is the average of response differences across the RCF leaves (colored squares) to which the patient is classified. **C**) Selection of individuals with significant τ(p), those with confidence intervals not containing the null hypothesis (green arrows), to produce a binary profile (bottom). **D**) Targeting of individuals in other studies using the binary profile of high benefit from treating the disease with B (e.g. brodalumab).

3 Results

Clinicians and psoriasis patients are challenged by choosing between competing biologics of similar efficacy but differing on pathways of action and safety issues. We used *teff* to determine a transcription profile in non-lesional skin at baseline that was associated with the highest differences in response between 12-week brodalumab treatment, an IL-17A antagonist, and placebo. We analyzed transcriptomic data of non-lesional skin from 3-phase 3 clinical trials (AMAGINE 1-2-3) publicly available at the GEO repository under accession number GSE117468. The data comprises clinical and gene transcription data of 92 individuals with reported European ancestry, with moderate-to-severe psoriasis, 67 treated with brodalumab (either 140mg or 210mg) and with 25 with placebo for a period of 12 weeks. Covariates included age, body mass index (BMI), and the response variable was the amelioration of psoriasis area-and-severity-index (PASI) between baseline and 12-week of treatment. See reproducible code for all the analyses in **Supplementary material**.

The differential gene expression at baseline of 48 genes was associated with the interaction between treatment, brodalumab or placebo, and amelioration of symptoms at week 12. The top gene was *NR4A2*, previously shown to regulate Th17-mediated autoimmune inflammation (Raveney *et al.*, 2013). The 48 genes were significantly enriched in different immune pathways. The expected individual brodalumab effect at baseline, estimated by *teff* across significant genes, was strongly explained by the observed %PASI improvement after treatment (logistic model Vs constant model χ^2=6.81, df=2, P=1.55×10^{-15}). For patients that followed placebo, there was no significant logistic relationship between the expected brodalumab effect at baseline and the finally observed %PASI improvement (χ^2=0.03, df=2, P=0.98). These observations demonstrated the forecasting power of the estimated brodalumab effect, as a baseline measure, to predict the actual observed response of patients that finally took brodalumab. We then used *teff* to target individuals in the entire Brodalumab trial and observed that those targeted by the profile of high brodalumab benefit had higher T cell abundance in non-lesional skin at baseline (Log-Fold Change=0.022, P=0.034), as inferred by *immunedeconv* (Sturm *et al.*, 2019). This is in line with the antagonist function of brodalumab on IL-17A, targeting the IL-23/Th17 axis (Kim and Krueger, 2017).

We finally targeted patients in an independent trial (GSE11903) of etanercept and assessed whether those with high expected benefit to brodalumab responded to etanercept, a TNF inhibitor. We observed that the profile of high brodalumab benefit and the T cell abundance in non-lesional skin at baseline strongly explained the negative response to etanercept (R^2=0.715, χ^2(2)=10.13, P=0.006). An early analysis of the transcriptomic data from this trial reported that etanercept responders were linked to the downregulation of IL-17 pathway genes (Zaba *et al.*, 2009). Here, we make the further observation that etanercept responders can already be linked at baseline to low T cell levels in non-lesional skin and high expected benefit to brodalumab. As such, our analyses suggest that the action of the TNF inhibitor, instead of targeting the IL-23/Th17 axis, is optimized in patients where the axis may not be strongly compromised. Consequently, response to etanercept could be increased at baseline by offering brodalumab to patients with high potential benefits and high levels of T cell count in non-lesional skin. Overall, we show that causal inference and targeting on transcriptomic data can inform the patient choice between these two biological drugs.

Funding

This research has received funding from the Spanish Ministry of Education, Innovation and Universities, the National Agency for Research and the Fund for Regional Development (RTI2018-100789-B-I00) and the Spanish Ministry of Science and Innovation through the "Centro de Excelencia Severo Ochoa 2019-2023" Program (CEX2018-000806-S).
Conflict of Interest: none declared.

References

Athey,S. *et al.* (2019) Generalized random forests. *Ann. Stat.*, **47**, 1179–1203.
Correa da Rosa,J. *et al.* (2017) Shrinking the Psoriasis Assessment Gap: Early Gene-Expression Profiling Accurately Predicts Response to Long-Term Treatment. *J. Invest. Dermatol.*, **137**, 305–312.
Kim,J. and Krueger,J.G. (2017) Highly Effective New Treatments for Psoriasis Target the IL-23/Type 17 T Cell Autoimmune Axis. *Annu. Rev. Med.*, **68**, 255–269.
Raveney,B.J.E. *et al.* (2013) Nuclear Receptor NR4A2 Orchestrates Th17 Cell-Mediated Autoimmune Inflammation via IL-21 Signalling. *PLoS One*, **8**.
Sturm,G. *et al.* (2019) Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. In, *Bioinformatics*. Oxford University Press, pp. i436–i445.
Tomalin,L.E. *et al.* (2020) Short-term transcriptional response to IL-17 receptor-A antagonism in the treatment of psoriasis. *J. Allergy Clin. Immunol.*, **145**, 922–932.
Wang,G. *et al.* (2020) Association of the Psoriatic Microenvironment With Treatment Response. *JAMA Dermatology*, **156**, 1057–1065.
Zaba,L.C. *et al.* (2009) Effective treatment of psoriasis with etanercept is linked to suppression of IL-17 signaling, not "immediate-response" TNF genes. *J. Allergy Clin. Immunol.*, **124**, 1022.