Response to the Lior Pachter's "nonsense" blog post

In his blog post dated February 11, 2014, Lior Pachter accuses us of committing scientific fraud in our Nature Biotechnology paper titled "Network deconvolution as a general method to distinguish direct dependencies in networks" (Vol 31, No 8, August 2013, pp. 726-733, published online on July 14, 2013). The blog post is repeating the same arguments brought forward in a commentary that Bray and Pachter submitted to Nature Biotechnology, which was reviewed and rejected by a panel of 4 referees. However, the blog post goes a step further and accuses us of scientific misconduct, fraud, and deceit.

These allegations are **defamatory, baseless, misleading, and factually incorrect**. We have already corresponded with Mr. Bray and Dr. Pachter at length, and responded to their questions over the past six months. We stand firmly by the scientific contributions of our work to the field of network science, and address the six allegations of misconduct below:

- Claim 1: "***the method used to obtain the results in the paper is completely different than the idealized version sold in the main text of the paper***". The paper clearly describes both the key matrix operation ("step 6" in Lior Pachter's blog post) which is shown in figure 1, and all the pre- and post-processing steps, which are all part of our method. There is nothing mischievous about including these pre- and post-processing steps that were clearly defined, well described, and implemented in the provided code.
- Claim 2: "***the method actually used has parameters that need to be set, yet no approach to setting them is provided***". It is unfortunate that so many methods in our field have parameters associated with them, but we are not the first method to have them. However, we do provide guidelines for setting these in the supplement.
- Claim 3: **"the authors appear to have deliberately tried to hide the existence of the parameters".** This claim is simply unfounded, as (a) we explicitly mention these parameter and give them names, (b) we provide options for setting them in the code, (c) we have dedicated sections of the paper describing the performance of the method under different parameters, (d) we include a figure showing the performance of our method under different settings of these parameters.
- Claim 4: "***the reason for covering up the existence of parameters is that the parameters were tuned to obtain the results***". Once more, we did not cover up the existence of parameters, and as we show below, our method is robust to these parameters across 27 regulatory networks in the DREAM5 network challenge (Appendix C).
- Claim 5: "***the results are not reproducible. The provided data and software is not enough to replicate even a single figure in the paper***". This claim is also unfounded: (a) All code and datasets have been publically available since publication of our manuscript; (b) in the News & Views about our paper, Alipanahi and Frey implemented our method and reported successful results back in August; (c) we have received several reports of successful applications of ND both using our provided code and with independent implementations; (d) we have since interacted with a scientist who has diligently reproduced each one of our key results. In addition, we now provide a supplementary page with detailed instructions for reproducing each one of our results at http://compbio.mit.edu/nd/ .
- Claim 6: "***the performance of the method [on a simulated dataset] is poor.***" As we discuss in Appendix D, the simulated dataset was tailored for partial correlation. In practice however, we have found partial correlation to perform poorly on real datasets, and we encourage Bray and Pachter to apply it to the DREAM5 benchmarks and to compare its results with our provided code.
- In addition to the six claims listed above, Dr. Pachter also points out that a correction to Supplement Figure S4 on August 26th 2013 was not fully documented. We apologize for the omission and provide additional details in an updated correction notice dated February 12, 2014. Importantly, we note that all corrections to the supplement were reviewed by the Nature Biotechnology editors and staff at all stages, were done in full transparency, and do not affect any of our conclusions. In particular, our method is robust to parameter choice as we show in Appendix C.

It unfortunately seems that Dr. Pachter's scrutiny of our work is personally motivated against senior author Manolis Kellis ever since 2007, and this perhaps explains the accusatory tone of his piece. However, we welcome the opportunity to provide this additional detail, and appreciate the scrutiny of our work that we hope will continue. We also hope the attention will prompt additional scientists to use our code, and increase the overall visibility and impact of our paper, which we firmly stand by.

Appendix A: Robustness to linear scaling (step 1)

The pre-processing step that Bray and Pachter criticize (step 1 in their description of our work) has no effect on the performance of ND. Mathematically, because our matrices are non-negative with min zero, a linear scaling of the network results in a similar scaling of its eigenvalues, which are normalized during eigenvalue scaling, canceling out the linear scaling of step 1. Practically, removing step 1 from the code has little effect on the performance of ND on the DREAM5 regulatory networks, as we show in Figure 1 below:

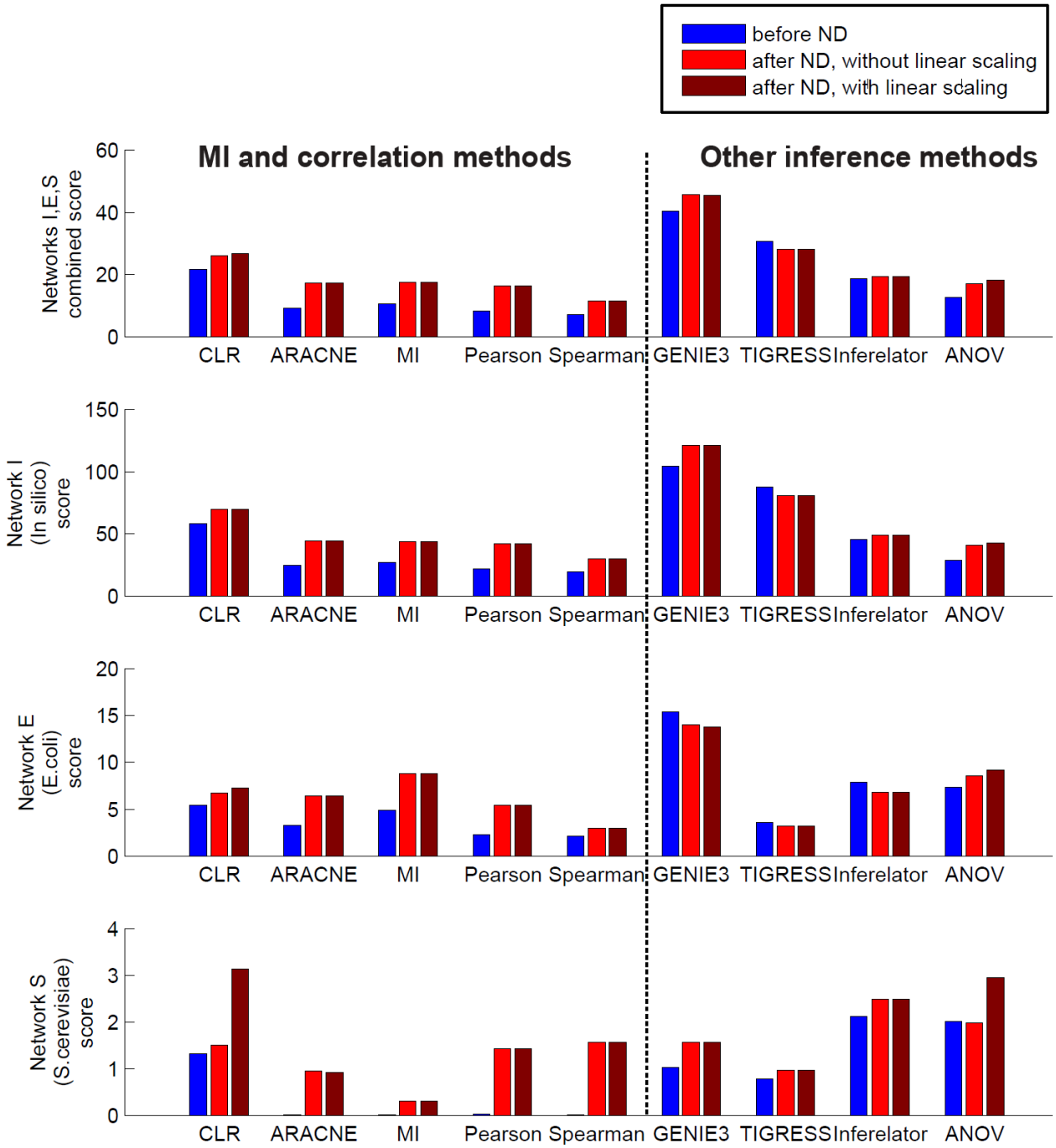

Figure 1: Performance of ND is robust to linear scaling ("step 1")

Appendix B: Importance of eigenvalue scaling (step 5)

The eigenvalue scaling (step 5) is essential both theoretically, to guarantee the convergence of the Taylor series, and practically, as performance decreases without it in practice. This step is clearly stated in both the main text of our paper and in the supplement, despite the claims by Bray and Pachter that it somehow was maliciously concealed from our description of the method (just search the manuscript for the word 'scaling'). Our empirical results confirm that this step is also necessary in practice, as ND without eigenvalue scaling is consistently performing worse, as we show in Figure 2 below:

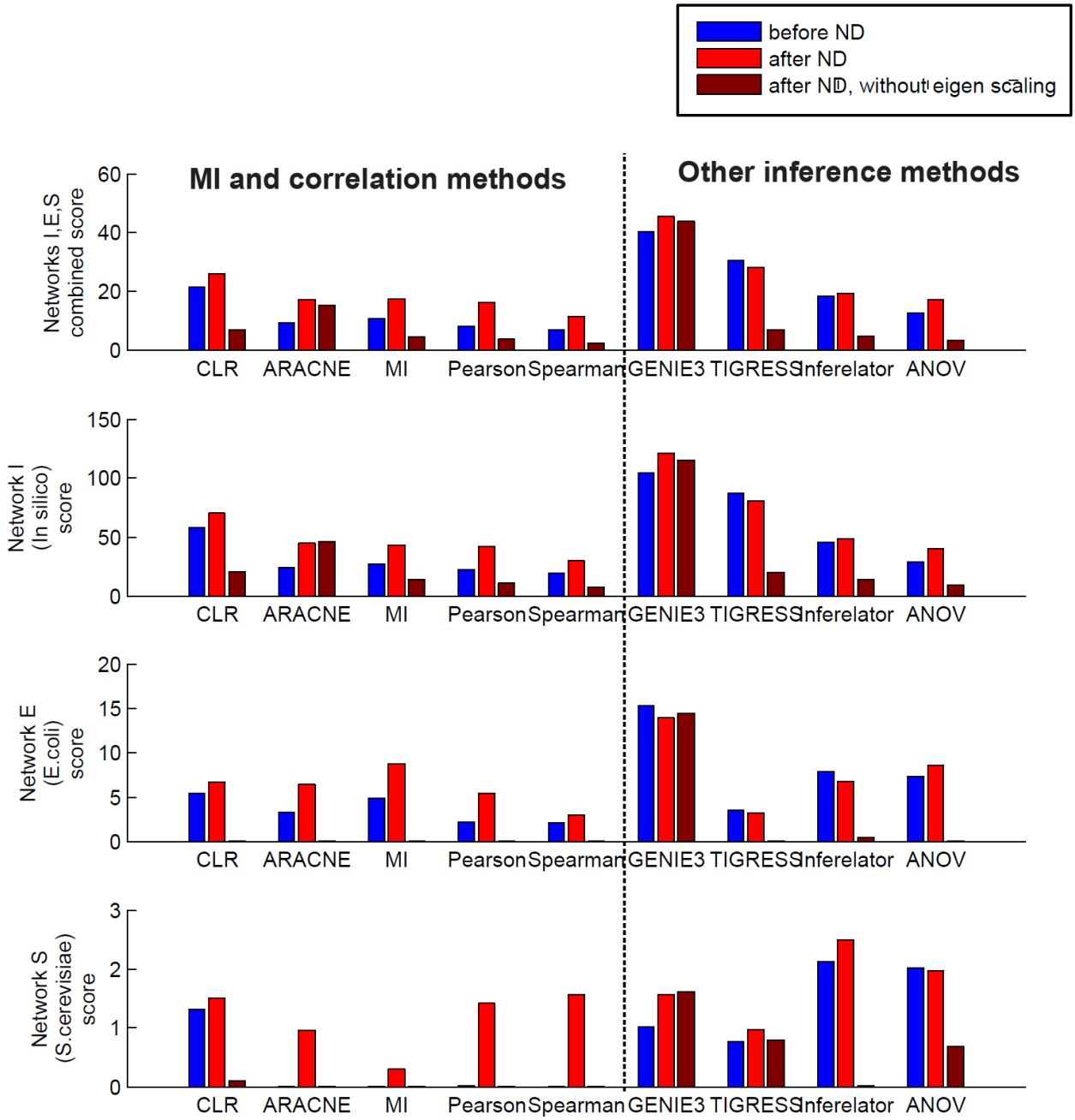

Figure 2: Eigenvalue scaling ("step 5") is necessary for network deconvolution.

Appendix C: Robustness to input parameters.

Bray and Pachter claim that "the reason for covering up the existence of parameters is that the parameters were tuned to obtain the results". Once again the claims are incorrect and unfounded. (1) First, we did not cover up the existence of the parameters. (2) Second, we used exactly the exact same set of parameters for all 27 tested regulatory networks in the DREAM challenge. (3) Third, we show here that our results are robust to the parameter values, using β=0.5 and β=0.99 for the DREAM5 network.

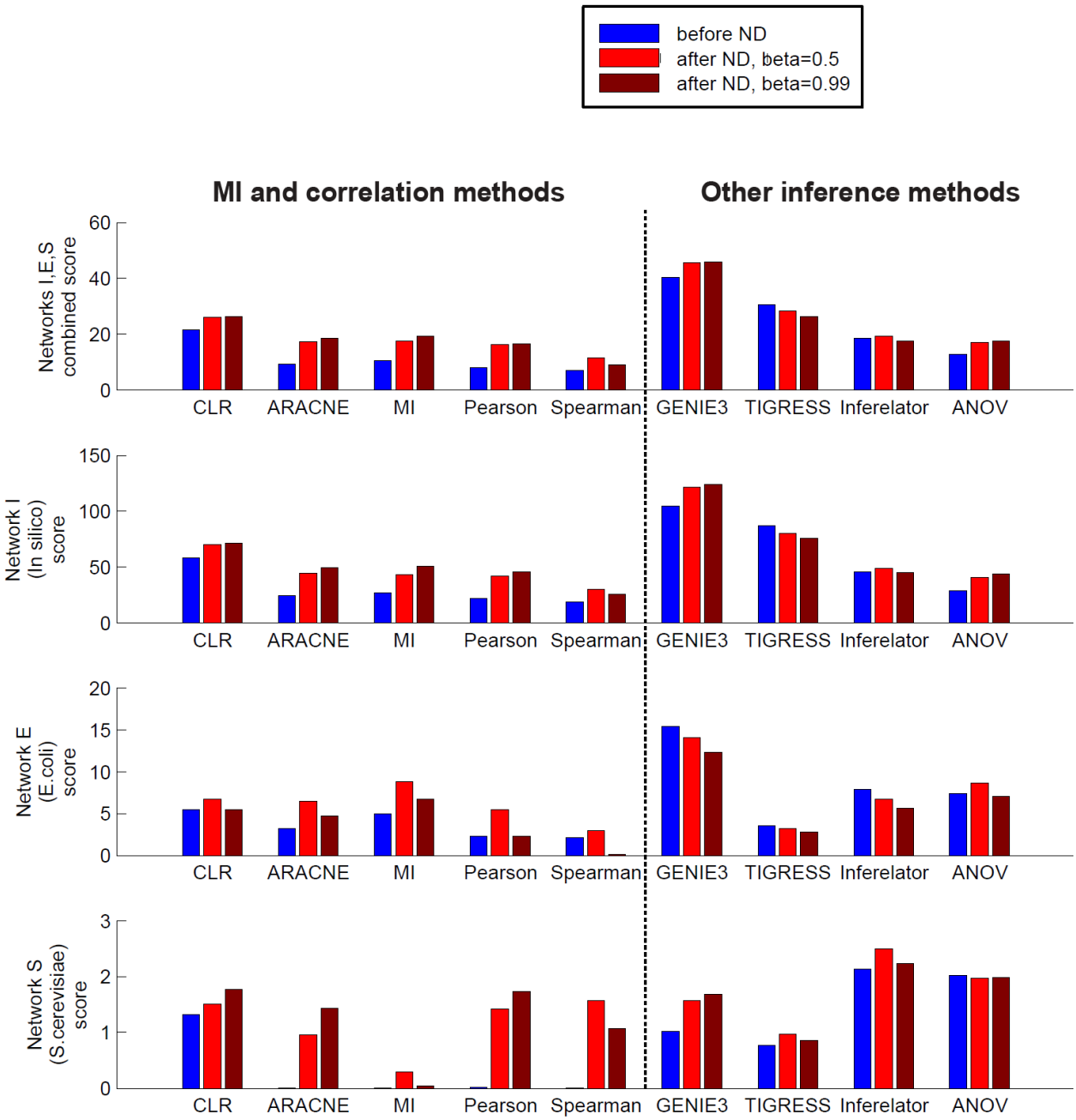

Figure 3: ND performance is robust to parameter choice.

Appendix D: Network Deconvolution vs. Partial Correlation.

Bray and Pachter compare network deconvolution to partial correlation using a test dataset built using a partial correlation model. In this very artificial setting, it is thus not surprising that partial correlation performs better, as it exactly matches the process that generated the data in the first place. To demonstrate the superiority of partial correlation, Bray and Pachter should test it on real datasets, such as the ones provided as part of the DREAM5 benchmarks. In our experience, partial correlation performed very poorly in practice, but we encourage Bray and Pachter to try it for themselves.