Comment on Feizi et al., Nat. Biotech. 31 (2013) by Nicolas Bray and Lior Pachter

The inference of direct interactions in networks using correlation estimates arguably began with the work of Seawall Wright in the 1920s on path coefficients for acyclic directed linear models[1,2]. He described a model for observed data G_{obs} in terms of direct and indirect effects that in modern language takes the form

$$G_{obs} = G_{indir} + G_{dir} \quad (1)$$

where

$$G_{indir} = G_{dir}^2 + G_{dir}^3 + \ldots \quad (2)$$

Later work on graphical models, especially covariance selection for Gaussian graphical models[3] can be seen as a continuation of Wright's program, a connection made explicit by Jones and West[4]. Echoing this work, Feizi *et al*.[5] claim to have developed a general method they state is "widely applicable in network science…across diverse disciplines". According to their Figure 1, their method, called "network deconvolution", computes a direct dependency matrix from an observed matrix by the formula

$$G_{dir} = G_{obs}(I+G_{obs})^{-1} \quad (3)$$

which is the solution for G_{dir} from (1). Feizi *et al.*'s justification for the general applicability of the model (1,2) and the deconvolution (3) is that it is correct for "linear time-invariant flow-preserving operators" but we fail to see how any interpretation of this phrase relates to the varied types of networks and matrices examined in the paper. However the applicability of (3) is ultimately irrelevant because it is not actually used anywhere in the paper outside of Figure 1.

Inspection of the code distributed in the Supplementary Material reveals the following procedure:

1) affinely map the entries of the matrix to lie between 0 and 1,
2) set the diagonal of the matrix to 0,
3) threshold the matrix, keeping only the largest α fraction of entries,
4) symmetrize the matrix,
5) scale the matrix so that the result of the next step will have maximum eigenvalue β,
6) apply formula (3),
7) affinely map between 0 and 1 again.

This procedure, which we call FK (for Feizi-Kellis according to the authors' description of contributions to the paper), differs not only from the description of the method in the paper and Online Methods, but also from the description in the Supplementary Materials. The mathematical interpretation of step 1, which is not mentioned anywhere in the Feizi *et al.* paper, is particularly unclear. While it is impossible to give a simple analytic formula for this procedure as a whole, using the Sherman-Morrison formula we found that when applied to a correlation matrix C, steps 1, 2, 4, 6 and 7 produce a matrix whose ijth entry is (up to an affine mapping)

$$P_{ij}+\Sigma_{kl}P_{ij}P_{kl}+m\Sigma_{kl}P_{ik}P_{jl} \quad (4)$$

where $P=C^{-1}$ and m is the minimum entry of C. Omitting step 1 results in P_{ii}, the inverse correlation matrix, so the effect of the mapping in this case is the addition of the final two terms of (4) whose possible meaning we were unable to decipher.

The FK procedure involves two parameters: α and β. The authors suggest setting β close to 1 in the Supplementary Material, but implement $\beta = 0.9$ as a default in the code (the default β was set to 1 when the paper was published but changed shortly thereafter). Given that there is no obvious way in which to choose β, nor is there a systematic approach to choosing the threshold parameter α, it seems curious that the settings used in the paper were[*]:

- DREAM5 challenge: $\beta = 0.5$ and $\alpha = 0.1$, and also omit step 4 in the above procedure.
- Protein network: $\beta = 0.99$ and $\alpha = 1$.
- Co-authorship: $\beta = 0.95$ and $\alpha = 1$.

The use of different parameters for different datasets is not disclosed in the paper and is troubling absent some rationale and methodology for setting them. Indeed, the DREAM5 challenge is blind specifically so that competitors cannot tune parameters in their predictions. Unfortunately, we were unable to replicate the results of the authors to check the performance of FK on the DREAM or protein datasets with other choices of parameters.

We therefore decided to examine how FK performs in one case for which Feizi *et al*. claim their method is optimal. They state "if the observed network is a covariance matrix of jointly Gaussian variables, ND infers direct interactions using global partial correlations". We can only imagine that the authors were referring here to a procedure consisting of only steps 2 and 6 in the above procedure and applied to a correlation (not covariance) matrix. However even then, this statement is false and neither equation (3) nor any other combination of steps 1-7 as used for the different results are equivalent to inference by partial correlation. Moreover, Figure 1 shows that FK performs significantly worse than partial correlation structural inference based on a James-Stein shrinkage estimation of the covariance matrix[6] that has been widely used for gene regulatory network analysis.

That FK does better than the most naïve method and yet worse than a method actually designed for this task is perhaps to be expected for a heuristic method that is based on a metaphor. It is easy to believe that in some contexts, FK will make things somewhat better. However a method that has mediocre performance in numerous settings is not of scientific value: every network problem arises in a specific domain and methods informed by those domains will yield superior results.

1. Wright, S. Correlation and Causation. *Journal of Agricultural Research* **20,** 557–585 (1921).
2. Wright, S. The Method of Path Coefficients. *The Annals of Mathematical Statistics* **5,** 161–215 (1934).
3. Dempster, A. P. Covariance Selection. *Biometrics* **28,** 157–175 (1972).
4. Jones, B. & West, M. Covariance decomposition in undirected Gaussian graphical models. *Biometrika* **92,** 779–786 (2005).
5. Feizi, S., Marbach, D., Médard, M. & Kellis, M. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nat Biotech* **31,** 726–733 (2013).
6. Schäfer, J. & Strimmer, K. A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology* **4,** (2005).

[*] Personal communication

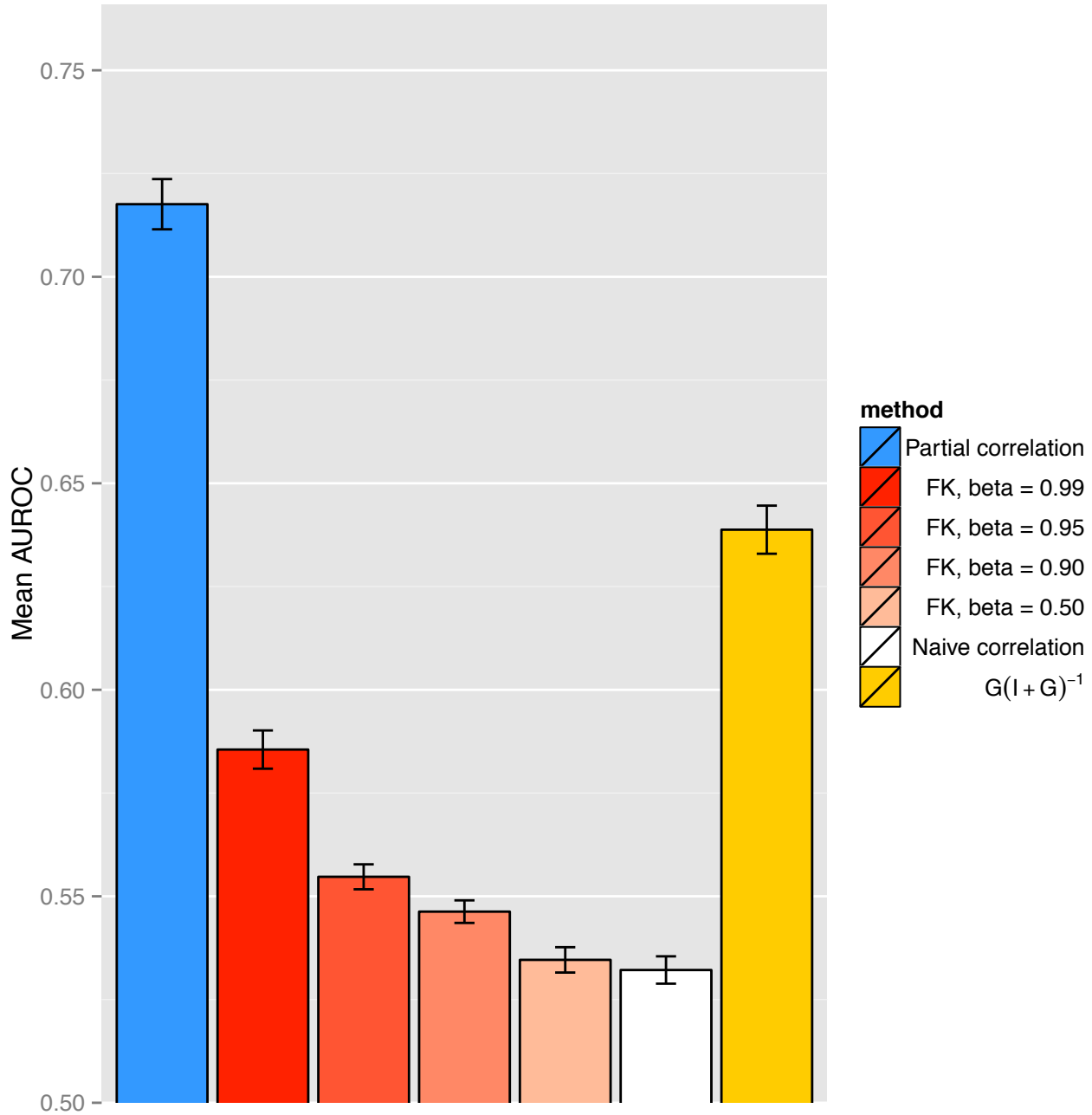

Figure 1: Comparison of the Feizi *et al.* method FK with the various beta parameters used in their paper to regularized partial correlation, the "geometric root" (equation (3)) and the naïve method of simply selecting the top entries of the correlation matrix. In order to test FK, we performed 40 simulations of sampling 500 observations from a random Gaussian graphical model with 1000 variables and an edge density of 5% to ensure the graph was connected yet sparse. Performance was assessed by comparing the ranking of the edges to the true edges present in the model and computing the area under the corresponding ROC curve. Because the FK procedure remaps its output to be between 0 and 1, we subtracted the median value of the output for a fair comparison.