IZKF Research Group Bioinformatics
Computational Biology Division

Institute for Biomedical Engineering
RWTH Aachen University Medical School

Aachen Institute for Advanced Study in Computational Engineering Science
RWTH Aachen University

Ivan G. Costa, PhD
Pauwelsstrasse 19 • 52074 Aachen, Germany
Tel.: +49-241-80 80270
www: http://www.costalab.org

Dear Dr. Rusk,

Thank you very much for your email on September 4, 2015 regarding our paper by Gusmao et al., "Addressing DNase-seq cleavage bias and residence time on computational footprinting" (NMETH-C24264A). We would like to thank you and the Reviewers for the thorough review of our paper and the constructive comments.

Following the Reviewers' and your requests, we have expanded our manuscript, which is now in analysis format. Moreover, we modified/adapted the focus of the manuscript as requested, which is now on the evaluation of computational footprinting methods. The manuscript includes now (i) a description of all relevant features of the evaluated methods, (ii) a detailed evaluation of the strategies for ranking footprints and (iii) a novel footprinting method (BinDNase). As before, we analyze the impact of DNase-seq artifacts, such as experimental bias and short residence time, on these methods.

As requested by Reviewer 2, our manuscript includes now a novel evaluation methodology. This methodology, referred to as FLR-Exp, is based on the quality assessment of footprints of a pair of cells with the fold change expression of the factor underlying the footprints. In contrast to the conventional evaluation methodology based on ChIP-seq data of transcription factors (TF), it only requires the gene expression data of the matching cells and TF motifs. Thus, FLR-Exp is more broadly applicable than the conventional TF ChIP-seq based evaluation. Interestingly, our experimental results show a clear concordance between the rankings of methods from both evaluation strategies. We have also improved the evaluation by the inclusion of further metrics, such as the area under the precision-recall curve (AUPR) and the area under the ROC curve (AUC) at distinct false discovery rates (FDR) on the TF ChIP-seq based evaluation.

We have also expanded and improved our analysis on experimental bias associated to DNase-seq experiments. As requested by Reviewer 2, we now clarify the differences between the two bias correction strategies explored in the manuscript: the “DNase-seq cleavage bias” and “DNase-seq experimental bias”. Our expanded analysis of cleavage bias on 61 DNase-seq experiments indicates that “DNase-seq experimental bias” is protocol-specific and differs from “DNase-seq cleavage bias”. Both strategies significantly improve footprint predictions in comparison to no bias correction. Finally, we have expanded our analysis on the association of CG content and bias correction strategies.

Altogether, we have addressed all requests by the Reviewers. Our manuscript presents the most comprehensive evaluation of computational footprinting methods up-to-date. We

provide a fair and reproducible benchmarking data set for evaluation of footprint predictions using two validation approaches. We provide statistics, predictions, data and software used in our evaluation for the community.

We would appreciate very much if you would consider our Analysis paper for publication in Nature Methods. In addition, we should be happy to answer any further question, which you might have.

With best regards,

Ivan G. Costa
Computational Biology Division
Institute for Biomedical Engineering
RWTH Aachen University