Cis-regulatory binding predicts gene expression across human tissues.

Keywords: gene – ChIP-Seq – transcription factors – multivariate regression –

Abstract: Cellular development requires the precise control of gene expression states. Transcription factors are involved in this regulatory process through their combinatorial binding with DNA. Regions along the genome with co-localized transcription factor binding sites may help determine gene expression levels. A handful of studies have shown that transcription factor binding information can be used to predict the expression levels of target genes, but only in model organisms or for just one cell type. Here we apply a novel approach to predict gene expression across various human tissue samples. Our additive model uses the expression and co-localized binding patterns of 14 transcription factors detected in K562 cells to predict target gene expression across many different tissues. A bootstrapping approach is applied to further assess how well co-localized transcription factors explain gene expression variation. We are able to explain as much as 64% of the gene expression variation in some tissues despite having limited information on transcription factors. Unlike traditional methods, our approach is also able to capture cooperation among transcription factors and their non-linear effect on target genes. We find that the contribution to target gene expression by each cis-regulatory module varies relative to its position from the transcription start site. Furthermore, larger cis-regulatory modules, ie., those with a higher number of co-localized transcription factors, explain more of their target gene's expression than smaller modules. Given that 45% of these detected cis-regulatory modules seem to have a non-linear effect on the expression of their target gene, this study highlights a need to better understand the divergent roles of gene regulators in different tissues.

Authors:

- Dennis Wang dennis.wang@mrc-bsu.cam.ac.uk United Kingdom MRC Biostatistics Unit, University of Cambridge
- Augusto Rendon ar506@cam.ac.uk United Kingdom Department of Haematology, University of Cambridge
- Lorenz Wernisch Lorenz.Wernisch@mrc-bsu.cam.ac.uk United Kingdom MRC Biostatistics Unit