OXFORD

Systems biology

RTNduals: an R/Bioconductor package for analysis of co-regulation and inference of *dual regulons*

Vinicius S. Chagas[1,†], Clarice S. Groeneveld[1,†], Kelin G. Oliveira[1,2], Sheyla Trefflich[3], Rodrigo C. de Almeida[4], Bruce A. J. Ponder[5], Kerstin B. Meyer[5,6], Steven J. M. Jones[7], A. Gordon Robertson[7,*,‡] and Mauro A. A. Castro[1,*,‡]

[1]Bioinformatics and Systems Biology Lab, Federal University of Paraná, Curitiba 81520-260, Brazil, [2]Department of Clinical Sciences, Section of Oncology and Pathology, Lund University, Lund 221 85, Sweden, [3]Graduate Program in Bioinformatics, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte 31270-901, Brazil, [4]Department of Biomedical Data Sciences, Molecular Epidemiology, Leiden University Medical Center, Leiden 2300 RC, The Netherlands, [5]Department of Oncology and Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 ORE, UK, [6]Wellcome Sanger Institute, Hinxton CB10 1SA, UK and [7]Canada's Michael Smith Genome Sciences Center, BC Cancer Agency, Vancouver V5Z 4S6, Canada

*To whom correspondence should be addressed.

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

[‡]The authors wish it to be known that, in their opinion, the last two authors should be regarded as Joint Last Authors.

Associate Editor: Janet Kelso

Abstract

Motivation: Transcription factors (TFs) are key regulators of gene expression, and can activate or repress multiple target genes, forming regulatory units, or regulons. Understanding downstream effects of these regulators includes evaluating how TFs cooperate or compete within regulatory networks. Here we present *RTNduals*, an R/Bioconductor package that implements a general method for analyzing pairs of regulons.

Results: *RTNduals* identifies a dual regulon when the number of targets shared between a pair of regulators is statistically significant. The package extends the *RTN* (Reconstruction of Transcriptional Networks) package, and uses *RTN* transcriptional networks to identify significant co-regulatory associations between regulons. The Supplementary Information reports two case studies for TFs using the METABRIC and TCGA breast cancer cohorts.

Availability and implementation: *RTNduals* is written in the *R* language, and is available from the Bioconductor project at http://bioconductor.org/packages/RTNduals/.

Contact: grobertson@bcgsc.ca or mauro.castro@ufpr.br

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Gene regulation in eukaryotes integrates a large number of interconnected regulatory influences. Some of the major contributors in gene regulation are transcription factors (TFs): proteins that can act as activators or repressors of gene expression, typically by binding to regulatory DNA regions and recruiting the transcriptional apparatus (Yamaguchi *et al.*, 2017). TFs are widely used in methods that reconstruct transcriptional networks, and algorithms that reconstruct
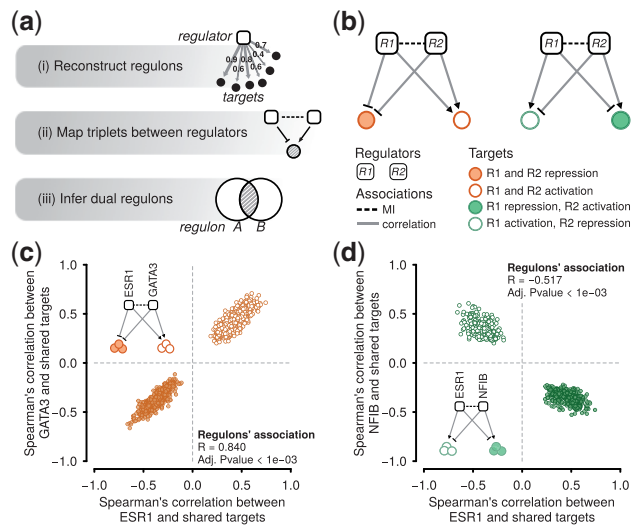
Fig. 1. Inference of dual regulons. (**a**) *RTNduals* computes dual regulons using: (i) MI between a regulator and targets; (ii) triplets consisting of pairs of regulators and a shared target; (iii) whether the number of shared targets is statistically significant. (**b**) Examples showing two associated regulators and two regulator-target triples. Left: an example in which the regulators co-operate by influencing shared target genes in the same direction (i.e. either co-activating or co-repressing the shared targets). Right: regulators compete by influencing shared target genes in opposite directions. (**c, d**) Distribution of correlation coefficients between regulators and shared targets in two example dual regulons, computed from the expression profiles of METABRIC breast cancer data, n = 997 for cohort 1 (Curtis *et al.*, 2012).

such networks consider both positive and negative target associations, consistent with mechanistic studies that have demonstrated the dual-function roles of TFs (Lubelsky and Shaul, 2019). In such a network, each regulator and its target genes form a regulatory unit, or regulon (Margolin *et al.*, 2006). Target genes can belong to multiple regulons, and regulators may co-operate and compete in influencing target gene expression.

In previous studies, we have used regulon activities to identify TFs associated with variant risk loci in breast cancer (Castro *et al.*, 2016), and to characterize differences between molecular subtypes in muscle-invasive bladder cancer (Robertson *et al.*, 2017). Because regulators can co-operate and compete, we anticipated that identifying pairs of regulons that share targets could be informative. Here, we report *RTNduals*, an R/Bioconductor package that automates the search for co-regulation between regulons, assessing all targets shared by pairs of regulators; when it identifies that a pair has more shared targets than expected by chance, which we assess by overlap and permutation analyses, it defines this pair as a *dual regulon*.

2 A method for identifying dual regulons

Figure 1a gives an overview of how *RTNduals* infers dual regulons. The package can take two types of data as input. The first type consists of a gene expression matrix (e.g. a cancer cohort's transcriptome) and, from prior biological information, a list that indicates which genes should be regarded as regulators. The second consists of a transcriptional regulatory network pre-computed by the *RTN* package (Castro *et al.*, 2016). The package architecture allows the input of different classes of regulators (e.g. TFs, miRNAs).

RTNduals uses three complementary statistics to identify dual regulons (Fig. 1a). (i) Targets are assigned to regulators based on

mutual information (MI), forming regulons. The statistical significance of the MI values is assessed by permutation and bootstrap analysis. Because regulators can target each other, associations between pairs of regulators are also identified. (ii) Triplets formed by two regulators and a shared target gene are identified, and the direction of regulation is determined by correlation analysis (e.g. Pearson or Spearman). (iii) A Fisher's exact test assesses the number of triplets shared between two regulators, and permutation analysis tests the statistical significance of the correlation between shared targets. The schematics in Figure 1b show the two cases that *RTNduals* identifies: regulator pairs (left) that co-operate, influencing shared target genes in the same direction (co-activation or co-repression), and (right) that compete, influencing targets in opposite directions. Figure 1c shows the distribution of Spearman correlations of targets shared between ESR1 and GATA3 regulons, indicating that these TFs either co-activate or co-repress their shared targets, while Figure 1d shows a contrasting case with ESR1 and NFIB regulons.

3 Case studies

RTNduals allows high-throughput screening for co-regulators and their shared targets. The Supplementary Information provides two detailed case studies that demonstrate the package's workflow, using tumor samples from breast cancer cohorts. The first study analyses a regulatory network generated by *RTN* from METABRIC microarray data (Curtis *et al.*, 2012), while the second case study shows how to prepare harmonized RNA-seq data from the National Cancer Institute's Genomic Data Commons (GDC) for analysis (TCGA, 2012).

References

Castro,M.A.A. *et al.* (2016) Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat. Genet.*, **48**, 12–21.

Curtis,C. *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, **486**, 346–352.

Lubelsky,Y. and Shaul,Y. (2019) Recruitment of the protein phosphatase-1 catalytic subunit to promoters by the dual-function transcription factor RFX1. *Biochem. Biophys. Res. Commun.*, **509**, 1015–1020.

Margolin,A.A. *et al.* (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, **7**, 1471–2105.

Robertson,A.G. *et al.* (2017) Comprehensive molecular characterization of muscle-invasive bladder cancer. *Cell*, **171**, 540–560.

The Cancer Genome Atlas Network. (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**, 61–70.

Yamaguchi,N. *et al.* (2017) Down-regulation of Forkhead box protein A1 (FOXA1) leads to cancer stem cell-like properties in tamoxifen-resistant breast cancer cells through induction of interleukin 6. *J. Biol. Chem.*, **292**, 8136–8148.