UNIVERSIDADE FEDERAL DO PARANÁ

VINICIUS SARAIVA CHAGAS

RTNduals: Ferramenta para analise de co-regulação entre regulons e inferencia de dual regulons

CURITIBA
2017

Vinícius de Saraiva Chagas

RTNduals: Ferramenta para análise de co-regulação entre *regulons* e inferência de *dual regulons*

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Bioinformática, no curso de Pós-Graduação em Bioinformática, Setor e Educação Profissional e Tecnológica da Universidade Federal do Paraná.

Orientador: Prof. Dr. Mauro Antonio Alves Castro

Coorientador: Prof. Dr. Rodrigo Almeida

Curitiba

2017

Chagas, Vinícius Saraiva de
C433 RTNduals: Ferramenta para análise de co-regulação entre regulons e inferência de dual regulons / Vinícius de Saraiva Chagas. – Curitiba, 2017
66 f.: il.

Orientador: Mauro Antônio Alves Castro
Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Educação Profissional e Tecnológica, Curso de Pós-Graduação em Bioinformática.

1. Redes regulatórias. 2. Co-regulação. 3. Fatores de transição. 4. MicroRNAs. 5. Moléculas - Bioquímica. I.Título. II. Castro, Mauro Antônio Alves. III. Universidade Federal do Paraná.

CDD 005.1

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA

Pós-Graduação em Bioinformática WWW.BIOINFO.UFPR.BR
E-mail: bioinfo@ufpr.br **Tel: 41 33614906**

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em BIOINFORMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de VINICIUS SARAIVA CHAGAS intitulada: "RTNduals: Ferramenta para analise de co-regulacao entre regulons e inferencia de dual regulons", após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa. A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 14 de Junho de 2017.

MAURO ANTONIO ALVES CASTRO
Presidente/Programa de Pós-graduação em Bioinformática – UFPR

JAQUELINE CARVALHO DE OLIVEIRA
Avaliadora Externa/Departamento de Genética – UFPR

ROBERTO TADEU RAITTZ
Avaliador Interno/Programa de Pós-graduação em Bioinformática – UFPR

Eu dedico este trabalho a minha família e meus amigos, mas dedico principalmente a IaIa. Este é para você, pois você sabia que eu chegaria aqui antes de mim.

Agradecimentos

Eu agradeço primeiramente ao meu orientador Prof. Dr. **Mauro Castro** pelos dois anos de aprendizado em alto nível e por todas as oportunidades oferecidas durante meu mestrado. Agradeço também ao meu co-orientador Dr. **Rodrigo Almeida** por ter dado ideias e dicas importantes durante todo o curso e ajudado a aperfeiçoar este trabalho.

Obrigado aos professores do programa de Pós-Graduação em bioinformática, **Roberto** e **Berenice** por me auxiliarem sempre que foi preciso e a queridíssima **Suzana** por me ajudar a organizar minha vida acadêmica.

Eu gostaria de agradecer a todos os colegas de laboratório, **Kelin**, **Clarice**, **Ani**, **Rodrigo** e em especial a parça de sempre **Sheyla** por terem aguentado meus comentários inoportunos e meu excelente gosto musical.

Eu agradeço também a todo o apoio que tive fora dos muros da universidade, obrigado **Gonzalo**, **Chito**, **Franco** e Prof. **Luis "Zão" Fávaro** por serem meus amigos e escutar todas minhas ladainhas e devaneios.

“Try not,
do or do not.
There is no try“
(Msc. Yoda)

Resumo

A regulação de expressão gênica é um fator chave nos processos biológicos e composta de uma extensa cascata de eventos que envolvem diversas moléculas e elementos. Entre estes elementos podemos destacar importantes reguladores tais como fatores de transcrição (FTs) e microRNAs (miRNAs), os quais podem ativar ou reprimir seus genes alvo e ainda cooperar ou competir em redes regulatórias. A complexidade dos efeitos destes reguladores requer o desenvolvimento de novos modelos capazes de integrar informação regulatória proveniente de diferentes tipos de dados. Embora exista uma variedade de ferramentas para o estudo de redes regulatórias, ainda existem deficiências metodológicas para integrar co-regulação entre reguladores e avaliar o efeito desta co-regulação em seus genes alvo. Neste trabalho nós apresentamos o *software RTNduals*, um pacote *R/Bioconductor* capaz de identificar *dual regulons*, um novo conceito que descreve pares de *regulons* cujos alvos em comum são afetados pelos dois reguladores. O pacote é uma extensão do software *RTN* (*Reconstruction of Transcriptional Networks*) e utiliza redes transcricionais para computar alvos compartilhados por dois reguladores e avaliar o efeito de ambos sobre os alvos compartilhados. O pacote *RTNduals* permite determinar se dois reguladores tem efeito similar ou oposto no conjunto de alvos compartilhados.

Palavras-chaves: Redes Regulatórias; Regulons; Co-regulação; Fatores de transcrição; microRNAs.

Abstract

The regulation of gene expression is a key factor in biological processes and it is composed of an extensive cascade of events involving several molecules and elements. We can highlight important regulators such as Transcription Factors (FTs) and microRNAs (miRNAs), which can activate or repress their target genes and still cooperate or compete in regulatory networks. The complexity of the effects among these regulators requires the development of new models capable of integrating regulatory information from different types of data. Although there are a variety of tools for the study of regulatory networks, there is still a methodological gap to integrate co-regulation between regulators and to evaluate the effect of this co-regulation on their target genes. In this work we present the *RTNduals*, an R/Bioconductor package capable of identifying *dual regulons*, which represent pairs of regulons whose common targets are affected by both regulators. The package is an extension of the Reconstruction of Transcriptional Networks (*RTN*) software and it uses *RTN*-generated transcriptional networks to test when pairs of regulators have a similar effect on their set of shared targets genes. With the set, *RTNduals* allows to determine whether two regulators have similar or opposite effect on their shared targets.

Key-words: Regulatory networks; Regulons; Co-regulation; Transcription Factors; microRNAs.

Lista de ilustrações

Sumário

Apresentação

Esta dissertação está estruturada em três partes. A ***Parte I*** apresenta uma introdução geral com ênfase em elementos regulatórios utilizados como modelos de estudo no desenvolvimento do *software RTNduals* (Fatores de transcrição e microRNAs) e na reconstrução de redes regulatórias utilizando a abordagem de reconstrução de *regulons*, assim como descreve motivos de rede existentes entre os reguladores, seguido dos objetivos gerais e especificos.

A ***Parte II*** apresenta cinco capítulos que descrevem em detalhes o pacote *RTNduals*. Esta parte contém o manuscrito do artigo científico que descreve o pacote, a vinheta do pacote com um fluxo de análises e um exemplo para a inferência de *dual regulons*, a página do pacote *RTNduals* no repositório *R/Bioconductor* e um estudo em parceria com outros colaboradores que mostra a importância das redes regulatórias no estudo de doenças complexas.

Por fim, a ***Parte III*** apresenta o fechamento do trabalho, contendo a discussão geral e conclusões com base nos resultados da *Parte II*.

Além destas três partes, a dissertação possui uma sessão de ***Anexos*** o conceito de *dual regulons* descritos neste trabalho e o manual das funcionalidades do pacote.

Parte I

Introdução Geral

1 Introdução

1.1 Elementos regulatórios da expressão gênica

A atuação precisa e bem organizada das etapas responsáveis pela expressão gênica é diretamente relacionada ao correto funcionamento dos processos biológicos celulares. No entanto, cada etapa possui uma complexa cadeia de eventos que envolvem diversos elementos regulatórios transcricionais e o efeito destes elementos sobre seus respectivos genes alvo (MASTON; EVANS; GREEN, 2006). A ação dos elementos regulatórios pode ocorrer em níveis distintos durante o processo de expressão do gene, podendo atuar como reguladores, transcricionais ou pós-transcricionais (THOMAS; CHIANG, 2006; KEENE, 2007). Dos elementos regulatórios citados anteriormente, os fatores de transcrição (FTs) e microRNAs (miRNAs) merecem destaque por sua participação regulatória em diversos processos celulares, entre os quais temos a proliferação, diferenciação e apoptose celular (CHUANG; CHIANG, 2014).

Os FTs fazem parte de um grupo de proteínas modulares que regulam pré-transcri cionalmente a expressão de um gene. Estas proteínas possuem dois domínios principais que são usados: (I) para se associar a sequências promotoras ou sequencias de *enhancer* do DNA e (II) para se associar a RNA-polimerase. Com isto os FTs são capazes de inibir a ação da RNA-polimerase e consequente transcrição do gene, ou, potencializar a ação da enzima na transcrição (SPITZ; FURLONG, 2012). Já os miRNAs, reguladores majoritariamente pós-transcricionais, consistem em pequenas moléculas (19 - 24 nucleotídeos) de RNA não-codificante. Estas moléculas encontram-se geralmente dentro da região 3'UTR (do inglês: untranslated region) dos seus genes alvo e são transcritas durante o processo de transcrição do gene. Estes RNAs não-codificantes modulam a transcrição de um gene, pelo pareamento, base a base, do miRNA com o mRNA do seu gene alvo, causando a degradação ou inibindo a tradução do mesmo (BARTEL, 2004; TOSCANO-GARIBAY; AQUINO-JARQUIN, 2014).

O alto potencial regulatório de FTs e miRNAs torna importante um melhor entendimento do efeito destes reguladores, de maneira sistêmica, na expressão gênica. Além disso, a super expressão ou repressão de alguns FTs ou miRNAs estão relacionados ao desenvolvimento de fenótipos de doenças multi-fatoriais como o câncer (NEBERT, 2002; PENG; CROCE, 2016).

O grande volume de informações disponíveis sobre estes tipos de moléculas gerou desafios para a criação de modelos organizacionais capazes de unir os dados de expressão genética, moléculas celulares e doenças (LAMB, 2006). Neste cenário a Biologia de Sistemas busca produzir modelos capazes de abranger toda a rede de ação de FTs e miRNAs, suas diferentes atuações nos genes, e a interação com outras moléculas (LEFEBVRE; RIECKHOF; CALIFANO, 2012).

1.2 Redes regulatórias e regulons

Modelos baseados em redes regulatórias oferecem insteressantes estratégias para elucidar complexos processos celulares utilizando diferentes tipos de dados. De acordo com BASSO et al. (2005), estes modelos consistem de uma intricada rede de interações moleculares entre moduladores, elementos regulatórios e conjunto de genes alvo.

Os modelos de redes regulatórias possuem diversos tipos de abordagens em sua construção, uma destas estratégias consiste na reconstrução de redes regulatórias transcricionais com base em *regulons* (MARGOLIN et al., 2006b). Nesta abordagem, os efeitos de um regulador sobre seus potenciais alvos são reconstruídos em uma unidade regulatória, ou *regulon*. Na **Figura 1** podemos observar o processo de reconstrução de um *regulon* utilizando matrizes de expressão (**Figura 1a**). A partir destas matrizes é computado o efeito do regulador (**Figura 1b**) e a direção deste efeito (**Figura 1c**) sobre seus potenciais alvos. Redes regulatórias baseadas em *regulons* podem ser utilizadas para visualizar e compreender a interação e os efeitos de diferentes elementos regulatórios com seus possíveis alvos, e assim, elucidar seus possíveis efeitos na expressão gênica (FLETCHER et al., 2013).

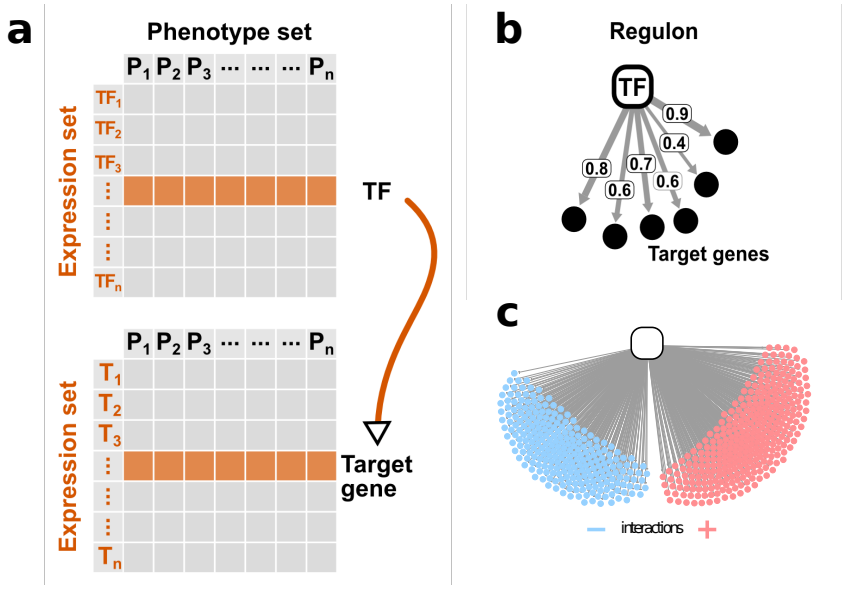

Figura 1: Representação do processo de reconstrução de unidades regulatórias (*regulons*). **(a)** Tipo de dado utilizado na reconstrução (matrizes de expressão).**(b)** Exemplo de *regulon* com valores de Informação Mútua. **(c)** Direção do efeito do regulador sobre seus potenciais alvos, em azul o regulador age negativamente sobre os alvos (repressão) e em vermelho positivamente (indução).

A arquitetura de uma rede regulatória é extremamente complexa, um exemplo dessa complexidade está ilustrado na **Figura 2**, a qual exibe as associações inferidas para uma rede regulatória centrada em FTs. Nesta rede, os FTs são representados por 2 reguladores (PTTG1 e E2F2) e cada *regulon* possui um FT como nó central. No seu entorno estão representados seus potenciais genes alvo e o efeito do regulador sobre o alvo.

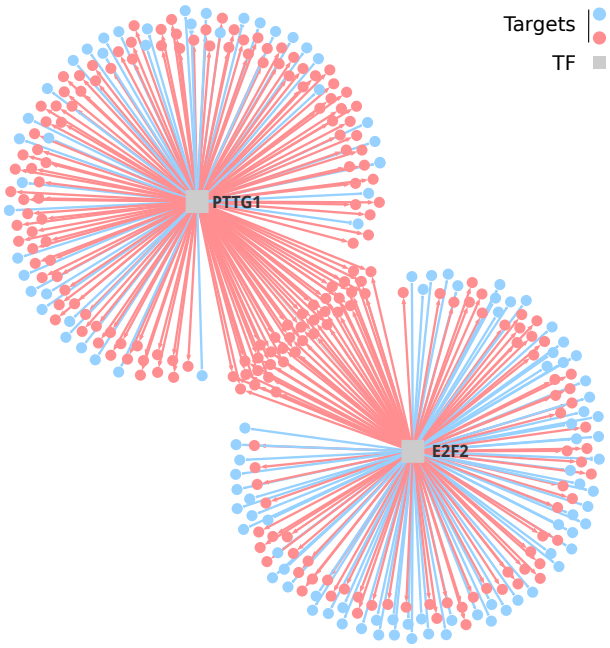

Figura 2: Complexidade de redes regulatórias. FTs são representados por quadrados e alvos por círculos. A cor vermelha indica indução da expressão do gene alvo pelo regulador e a cor azul representa repressão da expressão do gene alvo pelo regulador.

A complexidade de redes regulatórias ilustra os desafios para criação de ferramentas computacionais em Biologia de Sistemas. Modelos de estudo neste campo do conhecimento precisam ser capazes de integrar informações de diferentes tipos de reguladores, os quais atuam sobre o mesmo conjunto de alvos, tais como a ação co-regulatória entre FTs e miRNAs (LE et al., 2013; HAMED et al., 2015).

1.3 Motivos regulatórios e co-regulação

Fatores de transcrição (FTs) e microRNA (miRNAs) formam redes regulatórias extremamente complexas, compartilhando e co-regulando diversos alvos. Esta co-regulação tem produzido grande interesse na forma como estes elementos interagem entre si e agem em conjunto. A importância desta co-regulação é evidenciada pela correlação entre perturbações nas redes regulatórias destes dois elementos e fenótipos de algumas doenças de grande relevância epidemiológica (PENG et al., 2013; LIN et al., 2015). Além disso, a compreensão da co-regulação entre FTs e miRNAs tem forte apelo tanto para estudos do desenvolvimento celular, quanto para o desenvolvimento de novos fármacos.

A co-regulação entre estes reguladores pode ocorrer através de dois tipos de motivos regulatórios (*loops*): *feed-foward-loops* (FFLs) e *feedback-loops*(FBLs) (SHALGI et al., 2007; TSANG; ZHU; OUDENAARDEN, 2007), conforme ilustrado na **Figura 3**. Os FFLs representam motivos regulatórios em que existe ação entre os reguladores e ambos atuam sobre os mesmos alvos. Os FBLs representam motivos regulatórios em que existe apenas a ação entre os reguladores, mas estes não compartilham os mesmos alvos (INUI; MARTELLO; PICCOLO, 2010). Os FFLs entre FT e miRNA tem grande influência sobre o fenótipo celular, de acordo com INUI; MARTELLO; PICCOLO (2010) este tipo de *loop* atuaria como um "tampão molecular", minimizando o efeito do ruído causado por uma sinalização estocástica. Neste caso o FFL estabilizaria a tradução da proteína produto do sinal estocástico, ajustando a expressão do gene para uma direção contrária ao sinal.

Os FFLs podem ser divididos em diversos tipos. Estes *loops* podem ser classificados pelo seu regulador principal (**Figura 3a**), como: FT-FFL, miRNA-FFL e FFL composto. No tipo FT-FFL, o FT é o regulador principal, regulando tanto o miRNA quanto o gene alvo, no caso de um miRNA-FFL o miRNA é o regulador principal, regulando o FT e gene alvo. Os FFLs compostos ocorrem quando existe uma combinação de miRNA-FFL e FT-FFL, neste *loop* miRNA e FT se regulam entre si e ambos atuam sobre o mesmo alvo. Além da classificação baseada no regulador principal, os FFLs também podem ser classificados como coerentes (**Figura 3b**) ou incoerentes (**Figura 3c**) com base no efeito resultante do *loop* sobre seu gene alvo. No caso do FFL coerente tanto miRNA quanto FT atuarão sobre o gene e co-regulador para que tenham o mesmo resultado final, ativação ou repressão do gene. Enquanto no FFL incoerente, o efeito do regulador principal sobre o gene alvo é inconsistente quando comparado ao efeito do regulador principal sobre o co-regulador (ZHANG et al., 2013; ZHAO et al., 2016).

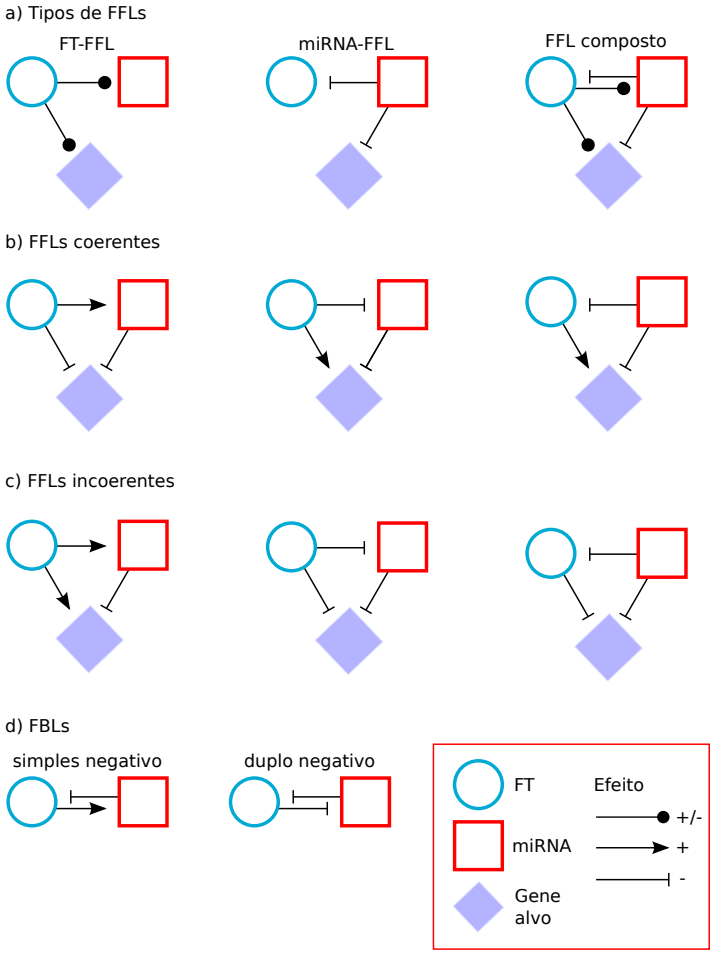

Figura 3: Tipos de *loops* encontrados entre FT, miRNA e gene alvo.**(a)**Tipos de FFLs de acordo com seu regulador principal. **(b)** FFLs coerentes, nestes tipos de *loops* o resultado final da co-regulação é coerente com o efeito que o regulador principal tem sobre seu gene alvo. Se o regulador principal inibe o alvo ele irá inibir o co-regulador que estiver induzindo a expressão e induzir a expressão de um co-regulador que também possui um efeito inibitório. O oposto ocorre nos FFLs incoerentes **(c)** aonde o efeito do regulador principal sobre o gene alvo é incoerente com o efeito sobre o regulador, neste caso se o regulador pirncipal inibe o gene alvo ele irá inibir o efeito de um co-regulador inibitório.**(d)** Tipos de FBLs. Adaptado de ZHANG et al. (2013).

A correlação entre a disfunção de mecanismos regulatórios e diversos tipos de doenças, como o câncer, tem revelado a importância de um melhor entedimento dos elementos regulatórios e suas formas de ação. Dessa forma, reguladores e suas unidades regulatórias são interessantes alvos terapêuticos, que em conjunto, criam a oportunidade de intervirmos em pontos chave do funcionamento celular (CASTRO et al., 2015). Portanto, quando se pensa em potenciais alvos para o tratamento de câncer, existe interesse na busca de moléculas com um amplo espectro de ação, como os FTs e miRNAs.

1.4 Justificativa

Em função do exposto acima, listamos a seguir algumas questões que motivaram os objetivos delineados neste trabalho:

i. Dificuldade do entendimento sistêmico da ação de reguladores de expressão gênica ***apesar*** da grande quantidade de dados moleculares referentes a estes reguladores;

ii. Falta de modelos que cubram o efeito de co-regulação de diferentes elementos regulatórios ***apesar*** da existência de modelos de redes regulatórias capazes de avaliar o efeito de um tipo de regulador;

iii. Pouca compreensão da co-regulação entre Fatores de Transcrição e microRNAs em modelos de redes transcricionais ***apesar*** da importância dos motivos regulatórios (*loops*) que estes elementos formam.

1.5 Objetivo Geral

Dada a importância da ação co-regulatória entre Fatores de Transcrição e microRNAs e as vantagens dos modelos de redes transcricionais para o estudo de regulação da expressão gênica, este trabalho teve como ***objetivo geral*** desenvolver uma ferramenta capaz de integrar redes transcricionais de diferentes reguladores e computar os possíveis *loops* entre Fatores de Transcrição e microRNAs.

1.6 Objetivos específicos

- Desenvolver uma ferramenta *R/Bioconductor* capaz de integrar redes regulatórias de diferentes tipos de elementos reguladores.
- Desenvolver um novo modelo para o estudo do efeito de co-regulação entre diferentes reguladores.
- Utilizar a ferramenta produzida para estudar os *loops* co-regulatórios entre Fatores de Transcrição e microRNAs

Parte II

RTNduals - pacote *R/Bioconductor*

2 Manuscrito

2.1 *RTNduals*: Um pacote *R/Biocondcutor* para análise de *loops* co-regulatórios e inferência de *dual regulons*

Este manuscrito descreve o pacote *RTNduals* e será submetido para publicação no periódico *Bioinformatics*.

Bioinformatics
doi.10.1093/bioinformatics/xxxxxx
Advance Access Publication Date: Day Month Year
Manuscript Category

OXFORD

Application Notes

RTNduals: An R/Bioconductor package for analysis of co-regulatory network loops and inference of *dual regulons*.

Vinicius S. Chagas [1,*], Clarice S. Groeneveld [1], Kelin G. Oliveira [1], Sheyla Trefflich [1], Rodrigo Coutinho de Almeida [2], Kerstin B. Meyer [3], A. Gordon Robertson [4] and Mauro A. A. Castro [1,*]

[1]Bioinformatics and Systems Biology Laboratory, Federal University of Paraná (UFPR), Polytechnic Center, Curitiba, Brazil.
[2]Department, Institution, City, Post Code, Country.
[3]Department, Institution, City, Post Code, Country.
[4]Department, Institution, City, Post Code, Country.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Transcription factors (TFs) and microRNAs (miRNAs) are key regulators of gene expression in many biological processes and can activate or repress their target genes and co-operate or compete on regulatory networks. The complexity of the downstream effects of these regulators requires development of new models capable of integrating regulatory information from a wide range of data types. Although a variety of tools are available to study regulatory networks, there still is a methodological gap to integrate co-regulation between regulators and to evaluate the effects of this co-regulation on target genes.
Results: Here we present *RTNduals*, an R/Bioconductor package to identify *dual regulons*, which represent pairs of regulons whose common targets are likely to be affected by both regulators. The package is an extension of the *RTN* (Reconstruction of Transcriptional Networks) package and uses the *RTN* transcriptional networks to test whether pairs of regulators have similar effects on their sets of target genes. Within the sets of shared targets, *RTNduals* allows to determine if two regulators have similar or opposing effects on their targets.
Availability: Our software is written in R language, and is available from the Bioconductor project at http://bioconductor.org/packages/RTNduals/
Contact: vinicius.chagas@ufpr.br and mauro.castro@ufpr.br
Supplementary information: Supplementary data (case studies) are available at *Bioinformatics* online.

1 Introduction

Gene regulation can involve a complex chain of events where a large number of regulatory elements might participate, such as Transcription Factors (TFs) and microRNAs (miRNAs), which are well studied key factors in several biological processes. TFs are modular proteins that can act as activator or repressor of gene expression by binding to the promoter region and recruiting the transcriptional apparatus [Lee and Young, 2000]. In contrast, miRNAs are small non-coding RNAs (around 19 to 25 nucleotides) that silence the expression of target genes by mRNA cleavage or translational repression Bartel [2004].

Although these regulators act at different levels (*i.e.* TFs are pretranscriptional regulators while miRNAs are posttranscriptional Chuang and Chiang 2014), both can co-operate or compete for the same targets in a regulatory network. The co-regulatory relationship between TFs and miRNAs has been demonstrated as an important factor in gene expression regulation, and the dysfunction of this co-operation has been associated with diverse diseases, including cancer Lin *et al.* [2015], Bracken *et al.* [2016]. The complexity of these relationships has been described by Zhang

et al. [2013] and classified into several types of network motifs between triplets of TFs, miRNAs and common targets. These motifs can form feed-forward loops (FFLs) and feedback loops (FBLs), either showing coherent or incoherent target effects. Despite the biological importance of these co-regulatory relationships, it is still challenging to develop models capable of integrating regulatory information from different types of regulators into a systems view and to evaluate the effects of this co-regulation on target genes. The available tools that study this kind of phenomena usually focus on computing the possible motifs between TFs, miRNAs and common targets [Le *et al.*, 2013, Hamed *et al.*, 2015]. However, they do not integrate co-regulation and evaluate the effects of upstream signaling on target genes. One possible approach is to use regulatory network models centered on the regulators. In these models, the downstream associations of a given regulator are reconstructed into a regulatory unit or regulon, and then, assessed in the context of gene perturbations [Fletcher *et al.*, 2013].

For example, the *RTN* (Reconstruction of Transcriptional Networks) is an R/Bioconductor package which computes transcriptional networks using regulons Castro *et al.* [2015]. This package uses Mutual Information (MI) to infer interactions between an individual regulator and its candidate target gene set (*i.e.* its regulon). For regulators that include steroid hormone receptors and transcription factors, the package has demonstrated that the activity of the regulon - as inferred by *RTN* - is more informative than the expression of the regulator itself . However, the *RTN* package does not cope with possible co-regulatory relationships between two types of regulators such as TFs and miRNAs. Here we provide a software that inherits *RTN* transcriptional networks and tests if, for any given pair of regulons, targets are likely to be affected by both regulators. *RTNduals* allows to analyze similar or different types of regulators (*e.g.* between TF-TF, miRNA-miRNA, TF-miRNA, TF-mRNA). The package tests the effects that two regulators have on their sets of shared target genes. Within the sets, *RTNduals* distinguishes when the two regulators have coherent or incoherent effects on the sets (REF). Pairs of regulons with significant similarities in their shared targets are named as *dual regulons*.

2 A method for discovering possible *dual regulons*

RTNduals can take as input two types of data in order to generate an MBR (Motifs Between Regulons) S4 class object. The first type consists of an expression matrix and two lists of regulators. The expression matrix is typically obtained from multiple samples (*e.g.* transcriptomes from a cancer cohort), while the lists of regulators represent some prior biological information indicating which genes in the expression matrix should be regarded as regulators. The package architecture allows the input of two classes of regulators such as TFs and microRNAs. In this case, the expression matrix should comprise mRNA and miRNA expression values. The other way to generate an MBR class is using two Transcriptional Network Inference (*TNI*)-*class* objects, pre-computed by the *RTN* package and used as the main data input.

RTNduals uses three complementary statistics to compute *dual regulons*: (i) Targets are assigned to regulons based on MI between the regulator and the target. The significance of the MI statistics is assessed by permutation and bootstrap analysis. The association between pairs of regulators is also identified in this step since regulators can target each other; (ii) Shared targets between any two regulons are identified and the similarity in regulation (which can be positive or negative) is assessed by correlation analysis (*e.g.* Pearson's or Spearman's correlation). Single network motifs are identified in this step. A motif is formed if two regulators that are linked in the network share a target gene, forming a triplet. A new motif if formed for each shared target, but only coherent motifs are considered; (iii) A test is carried out to determine if the correlation between the set of network motifs of any two regulons is higher than expected by chance. The **Figure 1a** shows an overview of the methodologies that are used by *RTNduals* in order to infer *dual regulons*. In **Figure 1b** we illustrate two *dual regulons*, each one with two regulatory motifs. In orange, the two regulators agree on the downstream effects (*i.e.* same directions), and an example of this agreement is showed in **Figure 1c**; whereas in green, they disagree (*i.e.* opposite directions) and **Figure 1d** shows an example of such disagreement.

Our method can be applied to any regulatory relationship based on regulons. For gene expression datasets, typical regulators might include transcription factors, miRNAs, eRNAs and long non-coding RNAs.

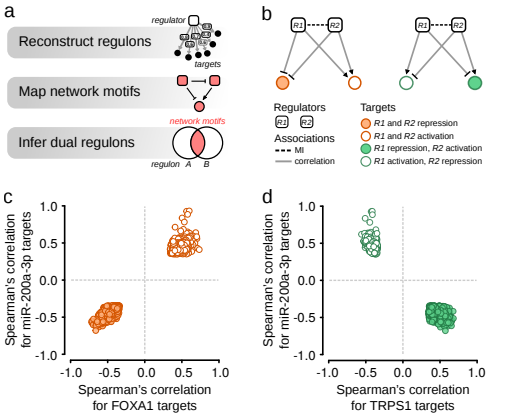

Fig. 1. RTNduals takes as input an expression matrix with two lists of regulators chosen by prior biological knowledge. The dual regulons are assessed by the package using the transcriptional network reconstruction methods provided by the RTN package. (a) RTNduals computes the dual regulons using: (i) MI between the regulator and the target; (ii) the association between pairs of regulators and their common targets; (iii) a hypergeometric test to evaluate if the set of network motifs of any two regulons is larger than expected by chance. (b) Two examples of two regulators with two network motifs each. In orange, an example in which the regulators agree at the downstream effects, in green the opposite case. (c, d) Scatter plots showing the association between regulons computed from TCGA BLCA data.

3 Core class and methods

The core S4 class in *RTNduals* is the MBR (Motifs Between Regulons). The S4 methods for this class cover the following steps:

Preprocessing: The S4 method 'mbrPreprocess' checks the consistency of the expression matrix and removes possible duplicated elements by selecting the most informative rows to compute MI. In this stage, the package also creates the regulatory networks using the two list of regulators.

Permutation and Bootstrap: The S4 method 'mbrPermutation' inherits the same algorithm implemented in the *RTN* package. In this process, MI is calculated between a regulator and the multiple targets in order to infer a regulon. Each regulon is computed with multiple hypothesis testing corrections. This step also removes unstable interactions by bootstrap analysis and creates a consensus bootstrap network.

Association: The S4 method 'mbrAssociation' takes the two transcriptional networks computed in the previous steps and enumerates all motifs between all regulons. The method retrieves the mutual information

between regulators and assesses the agreement between the predicted downstream effect using correlation analysis. A hypergeometric test is used to evaluate whether the common targets are potentially affected by both regulators in a number greater than expected by chance. A summary of the results can be accessed from the 'summary' slot using the 'mbrGet' function.

Acknowledgements

We thank for the postgraduate program in Bioinformatics at the Federal University of ParanÃ¡ and the collaboration of Dr. Gordon Robertson and Dra. Kerstin Meyer.

Funding: The Federal University of ParanÃ¡; CoordenaÃ§Ã£o de AperfeiÃ§oamento de Pessoal de Nivel Superior (CAPES);

References

Bartel, D. P. (2004). MicroRNAs. *Cell*, **116**(2), 281–297.

Bracken, C. P., Scott, H. S., and Goodall, G. J. (2016). A network-biology perspective of microRNA function and dysfunction in cancer. *Nature Reviews Genetics*, **17**(12), 719–732.

Castro, M. A. A., de Santiago, I., Campbell, T. M., Vaughn, C., Hickey, T. E., Ross, E., Tilley, W. D., Markowetz, F., Ponder, B. A. J., and Meyer, K. B. (2015). Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nature Genetics*, **48**(1), 12–21.

Chuang, T.-J. and Chiang, T.-W. (2014). Impacts of Pretranscriptional DNA Methylation Transcriptional Transcription Factor, and Posttranscriptional microRNA Regulations on Protein Evolutionary Rate. *Genome Biology and Evolution*, **6**(6), 1530–1541.

Fletcher, M. N. C., Castro, M. A. A., Wang, X., de Santiago, I., O'Reilly, M., Chin, S.-F., Rueda, O. M., Caldas, C., Ponder, B. A. J., Markowetz, F., and Meyer, K. B. (2013). Master regulators of FGFR2 signalling and breast cancer risk. *Nature Communications*, **4**.

Hamed, M., Spaniol, C., Nazarieh, M., and Helms, V. (2015). TFmiR: a web server for constructing and analyzing disease-specific transcription factor and miRNA co-regulatory networks. *Nucleic Acids Research*, **43**(W1), W283–W288.

Le, T. D., Liu, L., Liu, B., Tsykin, A., Goodall, G. J., Satou, K., and Li, J. (2013). Inferring microRNA and transcription factor regulatory networks in heterogeneous data. *BMC Bioinformatics*, **14**(1), 92.

Lee, T. I. and Young, R. A. (2000). Transcription of Eukaryotic Protein-Coding Genes. *Annual Review of Genetics*, **34**(1), 77–137.

Lin, Y., Zhang, Q., Zhang, H.-M., Liu, W., Liu, C.-J., Li, Q., and Guo, A.-Y. (2015). Transcription factor and miRNA co-regulatory network reveals shared and specific regulators in the development of B cell and T cell. *Scientific Reports*, **5**, 15215.

Zhang, H.-M., Kuang, S., Xiong, X., Gao, T., Liu, C., and Guo, A.-Y. (2013). Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. *Briefings in Bioinformatics*, **16**(1), 45–58.

3 Vinheta do pacote *RTNduals*

RTNduals: An R/Bioconductor package for analysis of co-regulatory network motifs and inference of 'dual regulons'.

Vinicius S. Chagas, Clarice S. Groeneveld, Kerstin B. Meyer, A. Gordon Robertson, Mauro A. A. Castro

2017-05-09

Abstract

Multiple regulators can co-operate or compete to control biological processes. RTNduals is a tool derived from the RTN (Reconstruction of Transcriptional Networks) package and searches for possible co-regulatory relationships between pairs of regulators. The tool infers "dual regulon" status, a new concept that tests whether pairs of regulators have similar effects on their sets of target genes (regulons). Members of a pair can be of similar types or different types (e.g. transcription factor (TF)-TF, miRNA-miRNA, TF-miRNA, TF-mRNA). RTN infers interactions between a list of regulators and candidate target gene sets (i.e. regulons) using Mutual Information (MI). The "dual regulon" extension identifies shared targets and assesses the similarity in regulation for these shared targets. RTNduals can identify regulators with shared or opposing effects on cellular phenotypes and can be applied to many different regulatory processes.

Contents

1 Overview

Recurrent patterns in biological networks have been termed network motifs (Bracken, Scott, and Goodall 2016), which may reflect critical roles in multiple biological processes; for example, regulatory loops between transcription factors and microRNAs (Zhang et al. 2015). **RTNduals** searches for targets shared between pairs of regulators, using regulatory networks generated by the **RTN package** (for details, please refer to the RTN documentation) (Castro et al. 2016). In such a network, each regulator has an associated set of gene targets (i.e. a regulon), and when we assess the shared targets in the regulons of a pair of regulators, we find that each shared target may be regulated in a positive or negative direction by a given regulator (i.e. pairs of regulators can either agree or disagree on the predicted downstream effects for a shared target gene). 'Dual regulons' represent regulon pairs whose common targets are likely to be affected by both regulators. The inference of 'dual regulons' requires three complementary statistics: (1) Targets are assigned to regulons based on MI between the regulator and the target. The significance of the MI statistics is assessed by permutation and bootstrap analysis. The association between pairs of regulators is also identified in this step, since regulators can target each other. (2) Shared targets between any two regulons are identified and the similarity in regulation (i.e. positive or

negative direction) is assessed by correlation analysis. Single network motifs are identified in this step consisting of two regulators and one common target. (3) A test is carried out to determine if the correlation between the set of network motifs of any two regulons is higher than would be expected by chance. The schematics in **Figure 1** show examples of 'dual regulons' with two network motifs each. In (a) the two regulators agree on the downstream effects (i.e. same directions), while in (b) they disagree (i.e. opposite directions). Our method can be applied to any regulatory relationship. For gene expression data sets typical regulators might include transcription factors, miRNAs, eRNAs and lncRNAs.

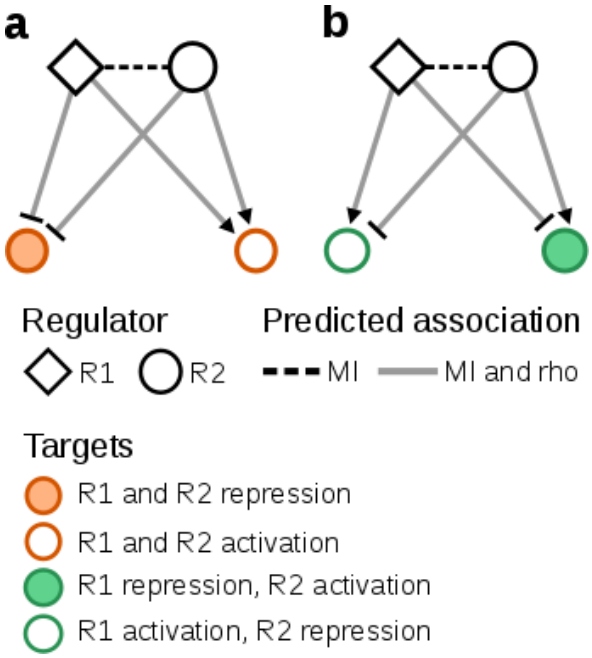

Figure 1 - Examples of regulators and predicted associations. This figure illustrates two 'dual regulons' and the respective downstream effects. (a) Regulators R1 and R2 agree in both activation (open circle) and repression (filled circle). (b) Regulators R1 and R2 disagree.

2 Quick Start

The **RTNduals** workflow starts with a preprocessing step that generates an **MBR-class** (Motifs Between Regulons) object from an expression matrix and two lists of regulators. The expression matrix is typically obtained from multiple samples (e.g. transcriptomes from a cancer cohort), while the lists of regulators represent some prior biological information indicating which genes in the expression matrix should be regarded as regulators. The input data can also deal with two classes of regulators; for example, genes and microRNAs. In this case, the expression matrix should comprise mRNA and miRNA expression values. Alternatively, the **MBR-class** object can be obtained by combining two **TNI-class** objects pre-computed in the **RTN** package.

2.1 Load datasets

This example provides the data required to generate an **MBR-class** object. The dataset **dt4rtn** is available from the **RTN** package and consists of an R list with 6 objects, 3 of which will be used in the subsequent analysis: (1) **gexp**, a named gene expression matrix with 250 samples (genes in rows, samples in cols), (2) **gexpIDs**, a data.frame with Probe-to-ENTREZ annotation, and (3) **tfs**, a named vector listing transcription factors (in this case, 149 TFs). While these datasets were extracted, pre-processed and size-reduced from Fletcher et al. (2013) and Curtis et al. (2012), they should be regarded as examples for demonstration purposes only.

```
##--- load package and prepare a dataset for demonstration
library(RTNduals)
```

```
data("dt4rtn", package = "RTN")
gexp <- dt4rtn$gexp
annot <- dt4rtn$gexpIDs
tfs1 <- dt4rtn$tfs[c("IRF8","IRF1","PRDM1","AFF3","E2F3")]
tfs2 <- dt4rtn$tfs[c("HCLS1","STAT4","STAT1","LMO4","ZNF552")]
```

2.2 Preprocessing

The **gexp** data matrix and the corresponding annotation are evaluated by the `mbrPreprocess` function in order to check the consistency of the input data. After this step it is generated a pre-processed **MBR-class** object whose status is updated to 'Preprocess [x]'.

```
##--- generate a pre-processed BR-class object
rmbr <- mbrPreprocess(gexp=gexp, regulatoryElements1=tfs1,
                      regulatoryElements2=tfs2, gexpIDs=annot, verbose=FALSE)
rmbr
## an MBR (Motifs Between Regulons) object:
## --status:
##  Preprocess Permutation   Bootstrap  DPI.filter Association
##         [x]         [ ]         [ ]         [ ]         [ ]
```

2.3 Run permutation analysis

The `mbrPermutation` method inherits the same algorithm implemented in the **RTN** package. This function takes the pre-processed **MBR-class** object and returns two regulatory networks that are inferred by mutual information analysis (with multiple hypothesis testing corrections). The results are included in the 'TNI' slots, which will be used in the subsequent steps of the pipeline.

```
##--- compute two regulatory networks
##--- (set nPermutations>=1000)
rmbr <- mbrPermutation(rmbr, nPermutations=100, verbose=FALSE)
rmbr
## an MBR (Motifs Between Regulons) object:
## --status:
##  Preprocess Permutation   Bootstrap  DPI.filter Association
##         [x]         [x]         [ ]         [ ]         [ ]
```

2.4 Run bootstrap analysis

In additional to the permutation analysis, the stability of the regulatory networks is assessed by bootstrapping using the `mbrBootstrap` function, which also inherits the same algorithm from the **RTN** package. Each 'TNI' slot of the **MBR-class** object is updated with a consensus bootstrap network.

```
##--- check the stability of the two regulatory networks
##--- (set nBootstrap>=100)
rmbr <- mbrBootstrap(rmbr, nBootstrap=10, verbose=FALSE)
rmbr
## an MBR (Motifs Between Regulons) object:
## --status:
##  Preprocess Permutation   Bootstrap  DPI.filter Association
##         [x]         [x]         [x]         [ ]         [ ]
```

2.5 Run DPI filter

In a given regulatory network each target can be linked to multiple regulators as a result of both direct and indirect interactions. The Data Processing Inequality (DPI) algorithm (P. Meyer, Lafitte, and Bontempi 2008) is used to remove the weakest interaction between two regulators and a common target. This step inherits the algorithm that is implemented in the **RTN** package, and it is optional for the analyses described in this workflow (the results of this step can be used to assess regulon activity in the **RTN** package).

```
##---apply DPI algorithm
rmbr <- mbrDpiFilter(rmbr, eps=0.05, verbose=FALSE)
rmbr
## an MBR (Motifs Between Regulons) object:
## --status:
##  Preprocess Permutation   Bootstrap  DPI.filter Association
##         [x]         [x]         [x]         [x]         [ ]
```

2.6 Run association analysis between regulons

The `mbrAssociation` method takes the two transcriptional networks computed in the previous steps and enumerates all motifs between all regulons. The method retrieves the mutual information between regulators and assesses the agreement between the predicted downstream effects using correlation analysis. A hypergeometric test is used to evaluate whether the common targets are potentially affected by both regulators in a number greater than expected by chance. Note that this example warns the user that only a few regulon pairs are being tested. This warning is to recall that the search space should ideally represent all possible combinations of a given class of regulators (for example, all nuclear receptors annotated for a given species).

```
##--- run the main RTNduals methods
rmbr <- mbrAssociation(rmbr, prob=0.75, verbose=FALSE)
## Warning: Only 25 regulon pair(s) is(are) being tested!
## Ideally, the search space should represent all possible
## combinations of a given class of regulators! For example,
## all nuclear receptors annotated for a given species.
```

A summary of the results can be accessed from the 'summary' slot using the `mbrGet` function.

```
##--- check summary
mbrGet(rmbr, what="summary")
## $MBR
## $MBR$Duals
##       testedDuals inferredDuals
## duals          25             7
##
##
## $TNIs
## $TNIs$TNI1
##          RE Targets Edges
## tnet.ref  5    2500  5728
## tnet.dpi  5    2500  5050
##
## $TNIs$TNI2
##          RE Targets Edges
## tnet.ref  5    2060  5285
## tnet.dpi  5    2060  4476
```

2.7 Rank 'dual regulons'

The `mbrDuals` method ranks all candidates using the correlation values computed in the `mbrAssociation` step, and returns the list of 'dual regulons'.

```
##--- run 'mbrDuals' and get results
rmbr <- mbrDuals(rmbr)
## -Sorting by the R value...
results <- mbrGet(rmbr, what="dualsInformation")
```

Also, when prior evidences are available this method can add a 'supplementaryTable' regarding the association between regulators. The 'supplementaryTable' is a 'data.frame' listing unique relationships between any two regulators used to compute the 'dual regulons' (please refer to the documentation for details on the input data format).

```
##--- here we build a 'toy' evidence table using the 'rnorm' function
supplementaryTable <- results[ ,c("Regulon1","Regulon2")]
supplementaryTable$ToyEvidence <- rnorm(nrow(results))
supplementaryTable
##             Regulon1 Regulon2 ToyEvidence
## IRF8~HCLS1      IRF8    HCLS1   1.2063271
## IRF8~STAT4      IRF8    STAT4  -0.1867745
## IRF1~STAT1      IRF1    STAT1  -1.9590200
## IRF1~STAT4      IRF1    STAT4  -0.9633900
## IRF1~HCLS1      IRF1    HCLS1  -0.7766609
## PRDM1~STAT4    PRDM1    STAT4  -0.5445293
## PRDM1~HCLS1    PRDM1    HCLS1  -0.9070042
```

```
##--- add supplementary evidences with the 'mbrDuals' function
rmbr <- mbrDuals(rmbr, supplementaryTable = supplementaryTable,
                 evidenceColname = "ToyEvidence", verbose = FALSE)
```

```
##--- check updated results
mbrGet(rmbr, what="dualsInformation")
##             Regulon1 Size.Regulon1 Regulon2 Size.Regulon2
## IRF8~HCLS1      IRF8          1062    HCLS1           969
## IRF8~STAT4      IRF8          1062    STAT4          1134
## IRF1~STAT1      IRF1           856    STAT1           907
## IRF1~STAT4      IRF1           856    STAT4          1134
## IRF1~HCLS1      IRF1           856    HCLS1           969
## PRDM1~STAT4    PRDM1          1086    STAT4          1134
## PRDM1~HCLS1    PRDM1          1086    HCLS1           969
##             Jaccard.coefficient Hypergeometric.Pvalue
## IRF8~HCLS1            0.7067227                     0
## IRF8~STAT4            0.6737805                     0
## IRF1~STAT1            0.6189164                     0
## IRF1~STAT4            0.5620094                     0
## IRF1~HCLS1            0.5678694                     0
## PRDM1~STAT4           0.5320911                     0
## PRDM1~HCLS1           0.5199704                     0
##             Hypergeometric.Adjusted.Pvalue        MI MI.Adjusted.Pvalue
## IRF8~HCLS1                               0 0.8481564              <0.01
## IRF8~STAT4                               0 0.6460023              <0.01
## IRF1~STAT1                               0 0.3989572              <0.01
## IRF1~STAT4                               0 0.3426818              <0.01
## IRF1~HCLS1                               0 0.3498595              <0.01
## PRDM1~STAT4                              0 0.2243653              <0.01
```

```
## PRDM1~HCLS1                              0 0.2805251              <0.01
##                      R Quantile ToyEvidence
## IRF8~HCLS1   0.8750054     1.00   1.2063271
## IRF8~STAT4   0.8568955     0.96  -0.1867745
## IRF1~STAT1   0.8086863     0.92  -1.9590200
## IRF1~STAT4   0.7835372     0.88  -0.9633900
## IRF1~HCLS1   0.7753173     0.84  -0.7766609
## PRDM1~STAT4 0.7539764     0.80  -0.5445293
## PRDM1~HCLS1 0.7400451     0.76  -0.9070042
```

2.8 Visualize shared cloud of targets

The package provides the `mbrPlotDuals` function to visualize 'dual regulon' and the respective motifs, which represent the shared cloud of targets. **Figure 2** shows two examples of 'dual regulons' for the schematics illustrated in **Figure 1**. In **(a)** the two regulators agree on the downstream effect (i.e. same directions), while in **(b)** they disagree (i.e. opposite directions).

```
duals <- mbrGet(rmbr, what="dualRegulons")
mbrPlotDuals(rmbr, names.duals = duals[1])
```

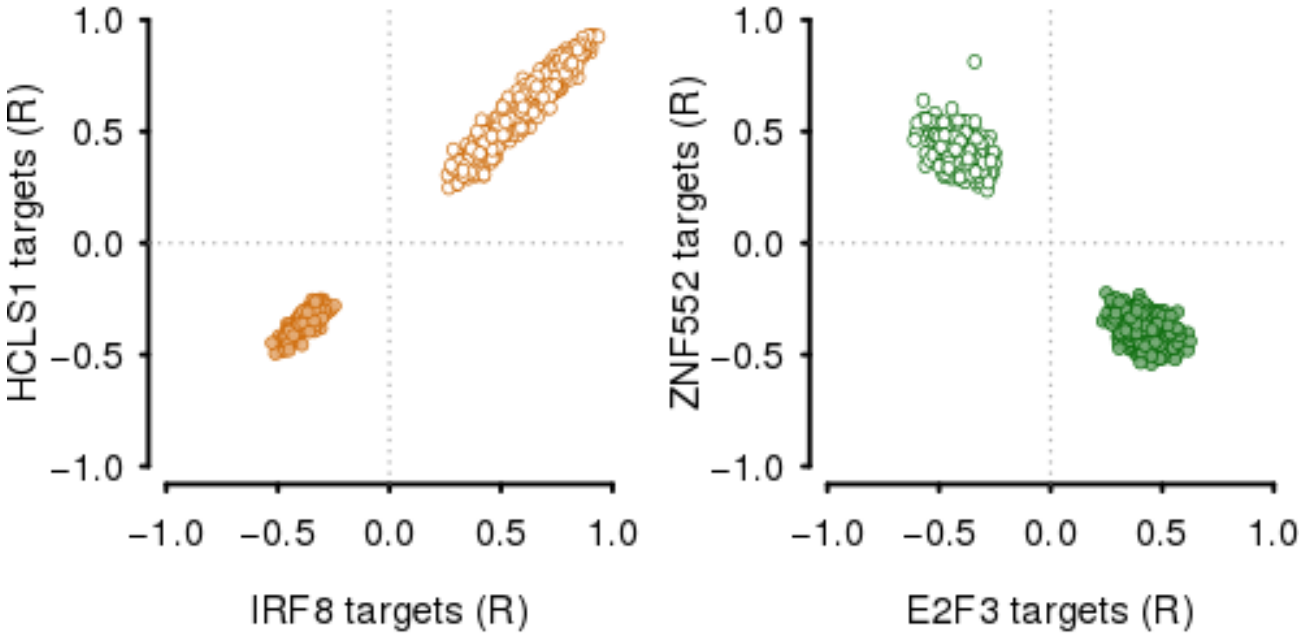

Figure 2. The 'mbrPlotDuals' function shows the shared cloud of targets. In (a) the regulators agree on the downstream effects, while in (b) they disagree. The colour pattern follows the schematics in Figure 1.

3 Session Info

```
sessionInfo()
## R version 3.3.3 (2017-03-06)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Debian GNU/Linux 8 (jessie)
##
## locale:
```

```
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] RTNduals_1.1.3  RTN_1.13.4      igraph_1.0.1    BiocStyle_2.2.1
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.8         plyr_1.8.4          nloptr_1.0.4
##  [4] bitops_1.0-6        tools_3.3.3         minet_3.32.0
##  [7] pvclust_2.0-0       digest_0.6.10       lme4_1.1-12
## [10] bit_1.1-12          evaluate_0.10       nlme_3.1-131
## [13] lattice_0.20-34     mgcv_1.8-16         ff_2.2-13
## [16] Matrix_1.2-8        yaml_2.1.14         parallel_3.3.3
## [19] SparseM_1.74        stringr_1.1.0       knitr_1.15.1
## [22] MatrixModels_0.4-1  S4Vectors_0.12.1    IRanges_2.8.1
## [25] stats4_3.3.3        rprojroot_1.1       nnet_7.3-12
## [28] grid_3.3.3          data.table_1.10.0   snow_0.4-2
## [31] fdrtool_1.2.15      XML_3.98-1.5        rmarkdown_1.3
## [34] minqa_1.2.4         limma_3.30.7        reshape2_1.4.2
## [37] corpcor_1.6.8       RedeR_1.22.0        car_2.1-4
## [40] magrittr_1.5        backports_1.0.4     htmltools_0.3.5
## [43] BiocGenerics_0.20.0 MASS_7.3-45         splines_3.3.3
## [46] pbkrtest_0.4-6      quantreg_5.29       stringi_1.1.2
## [49] RCurl_1.95-4.8
```

References

Bracken, Cameron P., Hamish S. Scott, and Gregory J. Goodall. 2016. "A Network-Biology Perspective of MicroRNA Function and Dysfunction in Cancer." *Nat Rev Genet* 17 (12). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 719–32. http://dx.doi.org/10.1038/nrg.2016.134.

Castro, Mauro, Ines de Santiago, Thomas Campbell, Courtney Vaughn, Theresa Hickey, Edith Ross, Wayne Tilley, Florian Markowetz, Bruce Ponder, and Kerstin Meyer. 2016. "Regulators of Genetic Risk of Breast Cancer Identified by Integrative Network Analysis." *Nature Genetics* 48 (1): 12–21. doi:10.1038/ng.3458.

Curtis, Christina, Sohrab P. Shah, Suet-Feung Chin, Gulisa Turashvili, Oscar M. Rueda, Mark J. Dunning, Doug Speed, et al. 2012. "The Genomic and Transcriptomic Architecture of 2,000 Breast Tumours Reveals Novel Subgroups." *Nature* 486: 346–52. doi:10.1038/nature10983.

Fletcher, Michael, Mauro Castro, Suet-Feung Chin, Oscar Rueda, Xin Wang, Carlos Caldas, Bruce Ponder, Florian Markowetz, and Kerstin Meyer. 2013. "Master Regulators of FGFR2 Signalling and Breast Cancer Risk." *Nature Communications* 4: 2464. doi:10.1038/ncomms3464.

Meyer, Patrick, Frederic Lafitte, and Gianluca Bontempi. 2008. "Minet: A R/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information." *BMC Bioinformatics* 9 (1): 461. doi:10.1186/1471-2105-9-461.

Zhang, Hong-Mei, Shuzhen Kuang, Xushen Xiong, Tianliuyun Gao, Chenglin Liu, and An-Yuan Guo. 2015. "Transcription

Factor and MicroRNA Co-Regulatory Loops: Important Regulatory Motifs in Biological Processes and Diseases." *Briefings in Bioinformatics* 16 (1): 45–58. doi:10.1093/bib/bbt085.

4 Repositório *Bioconductor*

O pacote *RTNduals* se encontra disponível para *download* na plataforma *Bioconductor* através do *link*:

http://bioconductor.org/packages/release/bioc/html/RTNduals.html

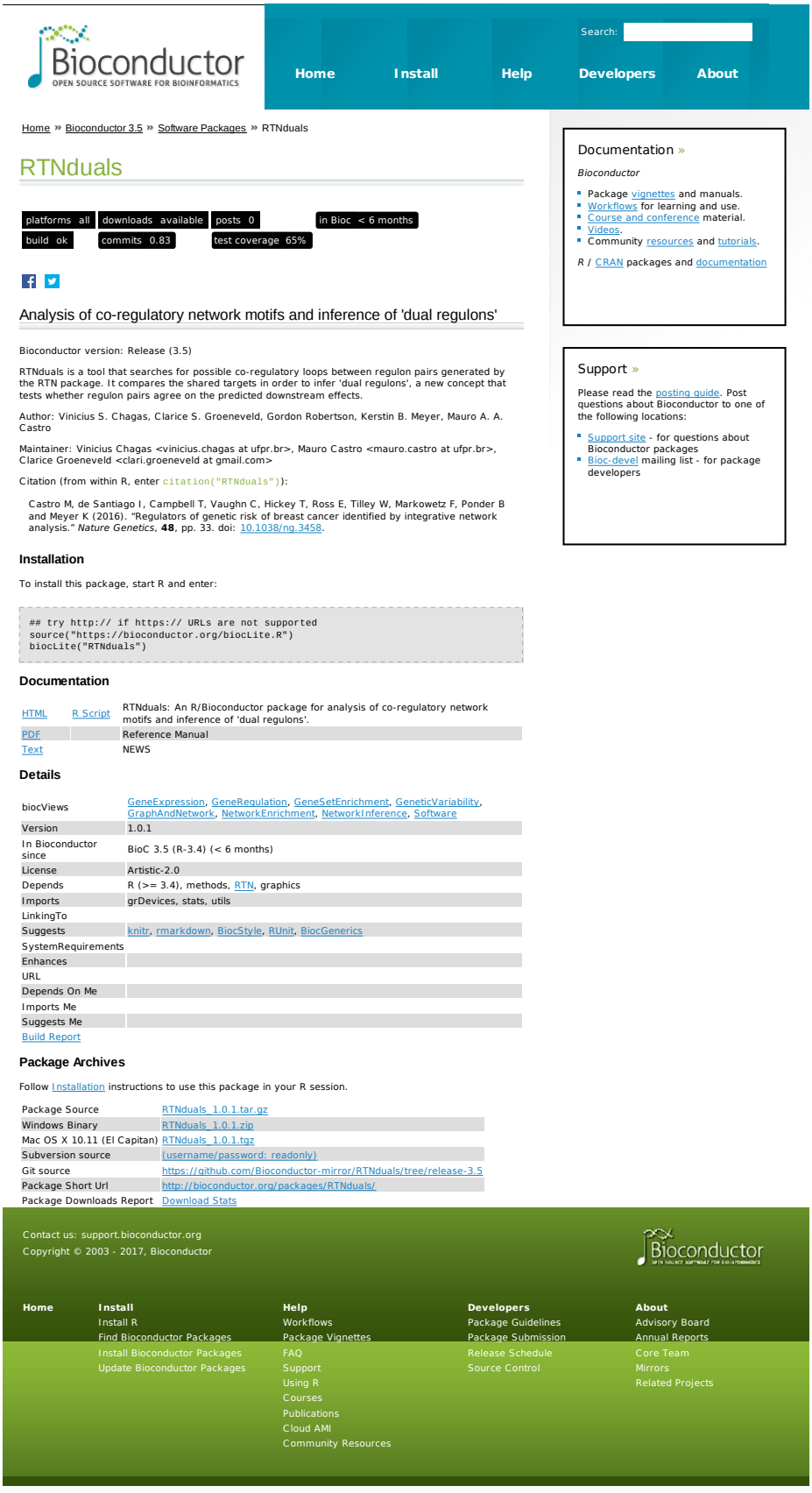

Figura 4: Página do pacote *RTNduals* na plataforma *Bioconductor*. Disponível em: http://bioconductor.org/packages/RTNduals/.

5 Manuscrito em colaboração

5.1 Manuscrito submetido ao periódico *Cell*, em parceria com o consórcio internacional *The Genome Cancer Atlas*

Este capítulo mostra parte de um estudo em colaboração com o consórcio *TCGA*, atualmente na segunda rodada de revisões no periódico *Cell*. Neste estudo usamos uma abordagem de redes regulatórias, reconstruindo *regulons* de Fatores de Transcrição e microRNAs para descrever o fenótipo de câncer de bexiga. Este estudo motivou o desenvolvimento de uma ferramenta capaz de integrar dois reguladores. O próximo passo será utilizar o pacote *RTNduals* para cruzar os resultados do pacote com as descobertas descritas neste estudo.

Cell

Comprehensive molecular characterization of muscle-invasive urothelial carcinoma

--Manuscript Draft--

Manuscript Number:	
Full Title:	Comprehensive molecular characterization of muscle-invasive urothelial carcinoma
Article Type:	Research Article
Keywords:	APOBEC mutation; basal mRNA subtype; CDKN2A; DNA methylation; lncRNA transcriptome; luminal mRNA subtype; muscle-invasive bladder cancer; microRNA; neoantigen; regulon
Corresponding Author:	Seth Lerner Houston, UNITED STATES
First Author:	Gordon Robertson
Order of Authors:	Gordon Robertson
	Jaegil Kim
	Hikmat Al-Ahmadie
	Joaquim Bellmunt
	Guangwu Guo
	Andrew Cherniack
	Toshinori Hinoue
	Peter Laird
	Katherine Hoadley
	Rehan Akbani
	Mauro Castro
	Ewan Gibb
	Rupa Kanchi
	Dmitry Gordenin
	Sachet Shukla
	Francisco Sanchez-Vega
	Donna Hansel
	Bogdan Czerniak
	Victor Reuter
	Xiaoping Su
	Benilton de Sa Carvalho
	Vinicius Chagas
	Karen Mungall
	Sara Sadeghi
	Chandra Pedamallu
	Yiling Lu
	Leszek Klimczak
	Jiexin Zhang

	Caleb Choo
	Akinyemi Ojesina
	Susan Bullman
	Kristen Leraas
	Tara Lichtenberg
	Catherine Wu
	Nicholaus Schultz
	Gad Getz
	Mathew Meyerson
	Gordon Mills
	David McConkey
	TCGA Research Network
	John Weinstein
	David Kwiatkowski
	Seth P Lerner
Abstract:	We report a comprehensive analysis of 412 muscle-invasive bladder cancers characterized by multiple TCGA analytical platforms. Fifty-eight genes were significantly mutated. Mutational load was mainly driven by APOBEC-mediated mutagenesis. Clustering by mutation signature identified a high-mutation subset with 89% 5-year survival. mRNA expression clustering refined prior clustering analyses, and identified a novel poor-survival 'neuronal' subtype that nevertheless lacked small cell or neuroendocrine histology. Clustering by mRNA, lncRNA, and miRNA expression converged to identify subsets with differential epithelial-mesenchymal transition, carcinoma-in-situ scores, histologic features, and survival. Our analyses identified 5 different subtypes of muscle-invasive bladder cancer that have different developmental mechanisms and that are likely to benefit from differential treatment approaches.
Opposed Reviewers:	Mattias Hoglund working in same specific area
	William Kim University of North Carolina Prior collaborator who pulled out of the present analytic group as he is using TCGA data for his immunology work
Suggested Reviewers:	Margaret Knowles University of Leeds m.a.knowles@leeds.ac.uk
	Francisco Real freal@cnio.es
	Nuria Malats nmalats@cnio.es
	Joshua Meeks joshua.meeks@northwestern.edu
	Elizabeth Plimack Elizabeth.Plimack@fccc.edu
	Francois Radvanyi francoisradvanyi@curie.fr

Comprehensive molecular characterization of muscle-invasive urothelial carcinoma

Authors

A. Gordon Robertson*[1], Jaegil Kim*[2], Hikmat Al-Ahmadie,[3] Joaquim Bellmunt,[4] Guangwu Guo,[5] Andrew D. Cherniack,[2] Toshinori Hinoue,[6] Peter W. Laird,[6] Katherine A. Hoadley,[7] Rehan Akbani,[8] Mauro A.A. Castro, [9] Ewan A. Gibb,[1] Rupa S. Kanchi,[8] Dmitry A. Gordenin,[10] Sachet A. Shukla,[5] Francisco Sanchez-Vega,[11] Donna E. Hansel,[12] Bogdan A. Czerniak,[13] Victor E. Reuter,[3] Xiaoping Su,[8] Benilton de Sa Carvalho,[14] Vinicius S. Chagas,[9] Karen L. Mungall,[1] Sara Sadeghi,[1] Chandra Sekhar Pedamallu,[2] Yiling Lu,[15] Leszek J. Klimczak,[16] Jiexin Zhang,[8] Caleb Choo,[1] Akinyemi I. Ojesina,[17] Susan Bullman,[2] Kristen M. Leraas,[18] Tara M. Lichtenberg,[18] Catherine J. Wu,[19] Nicholaus Schultz,[11] Gad Getz,[2] Matthew Meyerson,[20] Gordon B. Mills,[15] David J. McConkey,[21] TCGA Research Network, John N. Weinstein,[8] David J. Kwiatkowski[22] (co-corresponding author), Seth P. Lerner**[23]

*Co-First authors

Correspondence:

jweinste@mdanderson.org

dk@rics.bwh.harvard.edu

**slerner@bcm.edu (Lead)

Affiliations

[1]Canada's Michael Smith Genome Sciences Center, BC Cancer Agency, Vancouver, BC V5Z 4S6, Canada; [2]Cancer Program, The Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA; [3]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA; [4]PSMAR-IMIM Lab, Bladder Cancer Center, Department of Medicine, Dana-Farber Cancer Institute and Harvard University, Boston, MA 02215, USA; [5]Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard University, Boston, MA 02115,

USA; [6]Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI 49503, USA; [7]Department of Genetics, Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA; [8]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA; [9]Bioinformatics and Systems Biology Laboratory, Federal University of Paraná Polytechnic Center, Curitiba, PR CEP 80.060-000, Brazil; [10]Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, US National Institutes of Health, Research Triangle Park, NC 27709, USA; [11]Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA; [12]Department of Pathology, School of Medicine, University of California, San Diego, La Jolla, CA 92093, USA; [13]Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; [14]Biostatistics and Computational Biology Laboratory, Department of Statistics, University of Campinas, São Paulo, 13.083-859, Brazil; [15]Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; [16]Integrative Bioinformatics Support Group, National Institute of Environmental Health Sciences, US National Institutes of Health, Research Triangle Park, NC 27709, USA; [17]Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA; [18]Biospecimen Core Resource, The Research Institute at Nationwide Children's Hospital, Columbus, OH 43205, USA; [19]Department of Medical Oncology, Dana-Farber Cancer Institute; Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; [20]Pathology and Medical Oncology, Dana-Farber Cancer Institute and Harvard University, Boston, MA 02115, USA; [21]Greenberg Bladder Cancer Institute, Johns Hopkins University, Baltimore, MD 21218, USA; [22]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; [23]Scott Department of Urology, Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA

Summary

We report a comprehensive analysis of 412 muscle-invasive bladder cancers characterized by multiple TCGA analytical platforms. Fifty-eight genes were significantly mutated. Mutational load was mainly driven by APOBEC-mediated mutagenesis. Clustering by mutation signature identified a high-mutation subset with 89% 5-year survival. mRNA expression clustering refined prior clustering analyses, and identified a novel poor-survival 'neuronal' subtype that nevertheless lacked small cell or neuroendocrine histology. Clustering by mRNA, lncRNA, and miRNA expression converged to identify subsets with differential epithelial-mesenchymal transition, carcinoma-in-situ scores, histologic features, and survival. Our analyses identified 5 different subtypes of muscle-invasive bladder cancer that have different developmental mechanisms and that are likely to benefit from differential treatment approaches.

Keywords

APOBEC mutation, basal mRNA subtype, CDKN2A, DNA methylation, lncRNA transcriptome, luminal mRNA subtype, muscle-invasive bladder cancer, microRNA, neoantigen, regulon

Highlights

- High mutational load in bladder cancer is driven mainly by APOBEC-mediated mutagenesis
- Cancers with high-APOBEC and high-mutation load had an extraordinary 89% 5 year survival
- mRNA clustering identified a novel neuronal subtype with poor survival
- Convergent RNA clustering identified subtypes with different developmental mechanisms and distinct therapeutic potential

Introduction

Urothelial bladder cancer is a heterogeneous epithelial malignancy that presents most commonly as an exophytic tumor confined to the mucosa or lamina propria. However, 25% of

Parte III

Discussão Geral e Conclusões

6 Discussão Geral e Conclusões

A importância dos reguladores de expressão gênica no desenvolvimento e diferenciação celular já é bem estabelecida em diversos estudos (ANDERSON et al., 2002; KINGSLEY; BHAT, 2017). Destes reguladores, Fatores de Transcrição e microRNA ganham destaque devido sua ampla atuação reguladora e correlação com doenças complexas tais como o câncer (BHATLEKAR et al., 2017; GUO; BAO; YANG, 2017). Além disso, estes dois reguladores possuem conhecidos motivos regulatórios (*loops*) atuando em conjunto e co-regulando os mesmos alvos (ZHANG et al., 2013; ZHAO et al., 2016). Apesar da existência de ferramentas capazes de computar estes loops (LE et al., 2013; HAMED et al., 2015), ainda existe uma lacuna no entendimento dos efeitos desta co-regulação sobre seus alvos de maneira sistêmica.

Pensando na lacuna citada anteriormente, o presente trabalho desenvolveu um pacote na linguagem *R* (*Parte II*) capaz de computar os *loops* entre reguladores e genes alvo utilzando o modelo de redes regulatórias transcricionais baseadas em unidades regulatórias (*regulons*). A eficácia deste modelo já foi demonstrada por CASTRO et al. (2015) e FESSLER et al. (2016) utilizando a ferramenta *RTN* para reconstrução de redes regulatórias em câncer de mama. Além disso, o estudo descrito no *Capítulo 5* utiliza a reconstrução de *regulons* em redes regulatórias para estudar dois tipos de reguladores (Fatores de Transcrição e microRNAs) no fenótipo de câncer de bexiga. Neste estudo em parceria com o consórcio *TCGA* utilizamos *regulons* de Fatores de Transcrição e microRNAs para descrever o fenótipo de câncer de bexiga e como a condição (ativo ou inativo) destes *regulons* estaria associada com outras informações disponíveis. Porém, cada tipo de regulador foi usado separadamente no estudo, o que motivou o desenvolvimento de uma ferramenta capaz de integrar dois reguladores. A ferramenta *RTN* possui limitações para computar co-regulação entre *regulons* de diferentes redes. O pacote *RTNduals* utiliza redes regulatórias computadas pela ferramenta *RTN* e investiga a existência de *loops* entre os regulons de diferentes redes, com isso, a ferramenta é capaz de computar *dual regulons*,

um novo conceito desenvolvido durante este trabalho.

Os *dual regulons* seriam portanto a combinação dos conceitos discutidos anteriormente, ou seja, um modelo que engloba a descoberta de *loops* co-regulatórios entre *regulons* e o efeito desta co-regulação sobre os genes alvo. O modo como este modelo é computado, descrito em detalhes na *Parte II*, utiliza o conceito de *feed-foward-loops* (FFLs) e três diferentes tipos de estatísticas. Na primeira etapa a ferramenta computa Informação Mútua entre o regulador e seu potencial alvo de modo a reconstruir um *regulon*, o valor de significância desta Informação Mútua é calculado através de análise de permutação e *bootstrap* (MARGOLIN et al., 2006a; FLETCHER et al., 2013). Além disso, o algoritmo é capaz de mostrar qual a direção do efeito do regulador sobre seus alvos, podendo ser positivo (indução) ou negativo (repressão). O pacote *RTNduals* utiliza Informação Mútua para investigar possíveis *loops* entre os *regulons*, sendo que para ser considerado um *loop* deve existir Informação Mútua entre os dois reguladores e ambos devem atuar sobre um alvo compartilhado, conforme ilustrado na **Figura 5**. A concordância entre o efeito de ambos os reguladores (indução ou repressão) é então computada através de uma análise de correlação (*Spearman* ou *Pearson*), com isso podemos dizer se um possível *dual regulon* possui efeito concordante (**Figura 5a**) ou discordante (**Figura 5b**) sobre seus alvos compartilhados. Por fim, para ser considerado um *dual regulon* é realizado um teste hipergeométrico para mostrar se o número de alvos compartilhados entre os dois *regulons* é maior do que seria esperado ao acaso.

Os objetivos do trabalho, descritos na *Parte I*, são satisfeitos uma vez que o pacote *RTNduals* utiliza diferentes redes regulatórias transcricionais e produz resultados que integram as informações dos *regulons* de ambas as redes, conforme explicado no *Capítulo 3-4*. Esta integração tornou possível o modelo de *dual regulons*, que é capaz de rastrear os *loops* regulatórios entre os diferentes reguladores e avaliar seus efeitos sobre seus alvos compartilhados. O fato do pacote *RTNduals* utilizar o modelo de redes transcricionais produzido pelo pacote *RTN* permite que a ferramenta utilize diferentes tipos de dados de expressão (*RNA-seq*, *miRNA-seq*) como entrada, descrito no *Capítulo 3*. Além disso, o pacote *RTNduals* foi capaz de avaliar a co-regulação existente entre Fatores de Transcrição

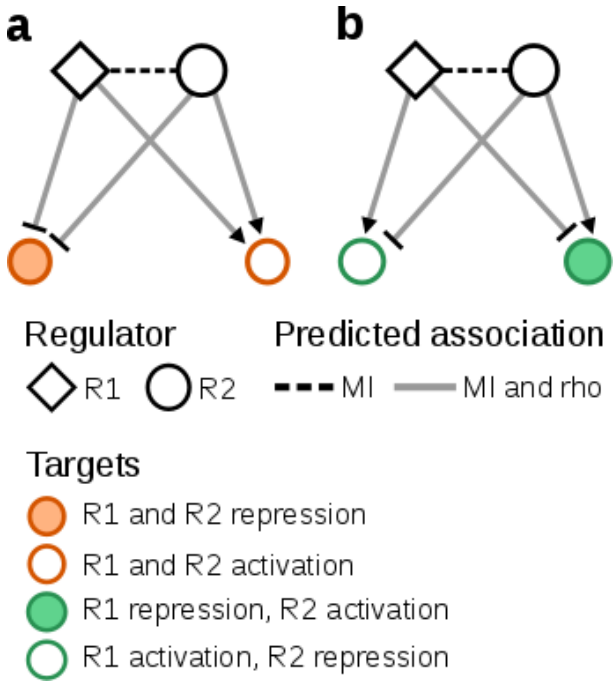

Figura 5: Exemplo de *dual regulons* e o efeito dos regulares sobre seus alvos compartilhados. **(a)** Ilustração de um *dual regulon* concordante, neste caso ambos reguladores R1 e R2 concordam na ativação (círculo vazio) e repressão (circulo preenchido) dos seus avlos. **(b)** Exemplo de *dual regulon* discordante, R1 e R2 tem efeito contrário em seus alvos compartilhados.

e microRNAs, conforme descrito no manuscrito do *Capítulo 2*.

O pacote *RTNduals* e o conceito de *dual regulons* abrem novas possibilidades para o estudo da co-regulação durante a expressão gênica. A versatilidade da ferramenta permite que reguladores de diferentes tipos sejam confrontados para verificar se existe um efeito co-regulador entre os elementos reguladores e qual a direção deste efeito sobre seus alvos. Assim, é possível verificar se existe contradição entre os reguladores ou se ambos atuam de forma semelhante sobre seus alvos compartilhados. Por fim, a co-regulação pode ser comparada entre diferentes fenótipos celulares, identificando semelhanças e diferenças entre grupos de *dual regulons*. O pacote *RTNduals* permitirá usar a informação de dois reguladores simultaneamente e portanto cruzar as informações dos *dual regulons* com os resultados do estudo descrito no *Capítulo 5*.

7 Perspectivas

O conceito de *dual regulons* abre discussão para novas questões nos efeitos de co-regulação entre reguladores, por exemplo, qual será o efeito da inativação de um dos reguladores em um *dual regulon* concordante? O outro regulador será capaz de compensar o efeito nos alvos? E em um *dual regulon* discordante? Haverá um aumento do efeito nos alvos pelo regulador ativo? As respostas para estas perguntas fazem parte dos próximos passos da utilização do pacote *RTNduals* em estudos futuros.

Referências

ANDERSON, M. K. et al. Definition of Regulatory Network Elements for T Cell Development by Perturbation Analysis with PU.1 and GATA-3. *Developmental Biology*, Elsevier BV, v. 246, n. 1, p. 103–121, jun 2002. Disponível em: <https://doi.org/10-.1006%2Fdbio.2002.0674>. Citado na página 44.

BARTEL, D. P. MicroRNAs. *Cell*, Elsevier BV, v. 116, n. 2, p. 281–297, jan 2004. Disponível em: <https://doi.org/10.1016%2Fs0092-8674%2804%2900045-5>. Citado na página 13.

BASSO, K. et al. Reverse engineering of regulatory networks in human B cells. *Nature Genetics*, Springer Nature, v. 37, n. 4, p. 382–390, mar 2005. Disponível em: <https://doi.org/10.1038%2Fng1532>. Citado na página 14.

BHATLEKAR, S. et al. Overexpression of HOXA4 and HOXA9 genes promotes self-renewal and contributes to colon cancer stem cell overpopulation. *Journal of Cellular Physiology*, Wiley-Blackwell, may 2017. Disponível em: <https://doi.org/10-.1002%2Fjcp.25981>. Citado na página 44.

CASTRO, M. A. A. et al. Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nature Genetics*, Springer Nature, v. 48, n. 1, p. 12–21, nov 2015. Disponível em: <https://doi.org/10.1038%2Fng.3458>. Citado 2 vezes nas páginas 18 e 44.

CHUANG, T.-J.; CHIANG, T.-W. Impacts of Pretranscriptional DNA Methylation Transcriptional Transcription Factor, and Posttranscriptional microRNA Regulations on Protein Evolutionary Rate. *Genome Biology and Evolution*, Oxford University Press (OUP), v. 6, n. 6, p. 1530–1541, jun 2014. Disponível em: <https://doi.org/10-.1093%2Fgbe%2Fevu124>. Citado na página 13.

FESSLER, E. et al. A multidimensional network approach reveals microRNAs as determinants of the mesenchymal colorectal cancer subtype. *Oncogene*, Springer Nature, v. 35, n. 46, p. 6026–6037, may 2016. Disponível em: <https://doi.org/10.1038%2Fonc-.2016.134>. Citado na página 44.

FLETCHER, M. N. C. et al. Master regulators of FGFR2 signalling and breast cancer risk. *Nature Communications*, Springer Nature, v. 4, sep 2013. Disponível em: <https://doi.org/10.1038%2Fncomms3464>. Citado 2 vezes nas páginas 14 e 45.

GUO, Y.; BAO, Y.; YANG, W. Regulatory miRNAs in Colorectal Carcinogenesis and Metastasis. *International Journal of Molecular Sciences*, MDPI AG, v. 18, n. 4, p. 890, apr 2017. Disponível em: <https://doi.org/10.3390%2Fijms18040890>. Citado na página 44.

HAMED, M. et al. TFmiR: a web server for constructing and analyzing disease-specific transcription factor and miRNA co-regulatory networks. *Nucleic Acids Research*, Oxford University Press (OUP), v. 43, n. W1, p. W283–W288, may 2015. Disponível em: <https://doi.org/10.1093%2Fnar%2Fgkv418>. Citado 2 vezes nas páginas 16 e 44.

INUI, M.; MARTELLO, G.; PICCOLO, S. MicroRNA control of signal transduction. *Nature Reviews Molecular Cell Biology*, Springer Nature, v. 11, n. 3, p. 264–275, mar 2010. Disponível em: <https://doi.org/10.1038%2Fnrm2868>. Citado na página 17.

KEENE, J. D. RNA regulons: coordination of post-transcriptional events. *Nature Reviews Genetics*, Springer Nature, v. 8, n. 7, p. 533–543, jul 2007. Disponível em: <https://doi.org/10.1038%2Fnrg2111>. Citado na página 13.

KINGSLEY, S. M. K.; BHAT, B. V. Role of MicroRNAs in the development and function of innate immune cells. *International Reviews of Immunology*, Informa UK Limited, p. 1–22, may 2017. Disponível em: <https://doi.org/10.1080%2F08830185.2017.1284212>. Citado na página 44.

LAMB, J. The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules Genes, and Disease. *Science*, American Association for the Advancement of Science (AAAS), v. 313, n. 5795, p. 1929–1935, sep 2006. Disponível em: <https://doi.org/10.1126%2Fscience.1132939>. Citado na página 14.

LE, T. D. et al. Inferring microRNA and transcription factor regulatory networks in heterogeneous data. *BMC Bioinformatics*, Springer Nature, v. 14, n. 1, p. 92, 2013. Disponível em: <https://doi.org/10.1186%2F1471-2105-14-92>. Citado 2 vezes nas páginas 16 e 44.

LEFEBVRE, C.; RIECKHOF, G.; CALIFANO, A. Reverse-engineering human regulatory networks. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, Wiley-Blackwell, v. 4, n. 4, p. 311–325, jan 2012. Disponível em: <https://doi.org/10-.1002%2Fwsbm.1159>. Citado na página 14.

LIN, Y. et al. Transcription factor and miRNA co-regulatory network reveals shared and specific regulators in the development of B cell and T cell. *Scientific Reports*, Springer Nature, v. 5, p. 15215, oct 2015. Disponível em: <https://doi.org/10.1038%2Fsrep15215>. Citado na página 16.

MARGOLIN, A. A. et al. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics*, Springer Nature, v. 7, n. Suppl 1, p. S7, 2006. Disponível em: <https://doi.org/10.1186%2F1471-2105-7-s1-s7>. Citado na página 45.

MARGOLIN, A. A. et al. Reverse engineering cellular networks. *Nature Protocols*, Springer Nature, v. 1, n. 2, p. 662–671, jul 2006. Disponível em: <https://doi.org/10-.1038%2Fnprot.2006.106>. Citado na página 14.

MASTON, G. A.; EVANS, S. K.; GREEN, M. R. Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics*, Annual Reviews, v. 7, n. 1, p. 29–59, sep 2006. Disponível em: <https://doi.org/10.1146%2Fannurev-.genom.7.080505.115623>. Citado na página 13.

NEBERT, D. W. Transcription factors and cancer: an overview. *Toxicology*, Elsevier BV, v. 181-182, p. 131–141, dec 2002. Disponível em: <https://doi.org/10.1016%2Fs0300-483x%2802%2900269-x>. Citado na página 14.

PENG, C. et al. Reconstruction and Analysis of Transcription Factor–miRNA Co-Regulatory Feed-Forward Loops in Human Cancers Using Filter-Wrapper Feature Selection. *PLoS ONE*, Public Library of Science (PLoS), v. 8, n. 10, p. e78197, oct 2013. Disponível em: <https://doi.org/10.1371%2Fjournal.pone.0078197>. Citado na página 16.

PENG, Y.; CROCE, C. M. The role of MicroRNAs in human cancer. *Signal Transduction and Targeted Therapy*, Springer Nature, v. 1, p. 15004, jan 2016. Disponível em: <https://doi.org/10.1038%2Fsigtrans.2015.4>. Citado na página 14.

SHALGI, R. et al. Global and Local Architecture of the Mammalian microRNA–Transcription Factor Regulatory Network. *PLoS Computational Biology*, Public Library of Science (PLoS), v. 3, n. 7, p. e131, 2007. Disponível em: <https://doi.org/10.1371%2Fjournal.pcbi.0030131>. Citado na página 17.

SPITZ, F.; FURLONG, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, Springer Nature, v. 13, n. 9, p. 613–626, aug 2012. Disponível em: <https://doi.org/10.1038%2Fnrg3207>. Citado na página 13.

THOMAS, M. C.; CHIANG, C.-M. The General Transcription Machinery and General Cofactors. *Critical Reviews in Biochemistry and Molecular Biology*, Informa UK Limited, v. 41, n. 3, p. 105–178, jan 2006. Disponível em: <https://doi.org/10-.1080%2F10409230600648736>. Citado na página 13.

TOSCANO-GARIBAY, J. D.; AQUINO-JARQUIN, G. Transcriptional regulation mechanism mediated by mirna–dna•dna triplex structure stabilized by argonaute. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, v. 1839, n. 11, p. 1079 – 1083, 2014. ISSN 1874-9399. Disponível em: <http://www.sciencedirect.com-/science/article/pii/S1874939914002041>. Citado na página 13.

TSANG, J.; ZHU, J.; OUDENAARDEN, A. van. MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals. *Molecular Cell*, Elsevier BV, v. 26, n. 5, p. 753–767, jun 2007. Disponível em: <https://doi.org/10.1016%2Fj-.molcel.2007.05.018>. Citado na página 17.

ZHANG, H.-M. et al. Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. *Briefings in Bioinformatics*, Oxford University Press (OUP), v. 16, n. 1, p. 45–58, dec 2013. Disponível em: <https://doi.org/10.1093%2Fbib%2Fbbt085>. Citado 4 vezes nas páginas 8, 17, 18 e 44.

ZHAO, Q. et al. Effect of Dynamic Interaction between microRNA and Transcription Factor on Gene Expression. *BioMed Research International*, Hindawi Publishing Corporation, v. 2016, p. 1–10, 2016. Disponível em: <https://doi.org/10.1155%2F2016%2F2676282>. Citado 2 vezes nas páginas 17 e 44.

Anexos

ANEXO A – Manual do pacote *RTNduals*

O **Anexo A** apresenta o manual de funcionalidades do pacote *RTNduals*. Neste capítulo encontramos a descrição da classe *MBR*, os métodos referentes a esta classe, os parâmtros utilizados em cada um dos métodos e as saídas resultantes dos métodos.

Package 'RTNduals'

May 3, 2017

Type Package

Title Analysis of co-regulatory network motifs and inference of 'dual regulons'

Version 1.0.1

Author Vinicius S. Chagas, Clarice S. Groeneveld, Gordon Robertson, Kerstin B. Meyer, Mauro A. A. Castro

Maintainer Vinicius Chagas <vinicius.chagas@ufpr.br>, Mauro Castro <mauro.castro@ufpr.br>, Clarice Groeneveld <clari.groeneveld@gmail.com>

Depends R (>= 3.4), methods, RTN, graphics

Imports grDevices, stats, utils

Suggests knitr, rmarkdown, BiocStyle, RUnit, BiocGenerics

Description RTNduals is a tool that searches for possible co-regulatory loops between regulon pairs generated by the RTN package. It compares the shared targets in order to infer 'dual regulons', a new concept that tests whether regulon pairs agree on the predicted downstream effects.

License Artistic-2.0

biocViews GeneRegulation, GeneExpression, GeneticVariability, GeneSetEnrichment, NetworkEnrichment, NetworkInference, GraphAndNetwork

LazyData TRUE

VignetteBuilder knitr

RoxygenNote 6.0.1

NeedsCompilation no

R topics documented:

`RTNduals-package` *RTNduals: An R/Bioconductor package for analysis of co-regulatory network motifs and inference of 'dual regulons'.*

Description

RTNduals is a tool that searches for possible co-regulatory loops between regulon pairs generated by the RTN package. It compares the shared targets in order to infer "dual regulons", a new concept that tests whether regulon pairs agree on the predicted downstream effects.

Details

Package:	RTNduals
Type:	Package
Depends:	R (>= 3.4.0), methods, RTN
Imports:	grDevices, stats, utils
Suggests:	knitr, rmarkdown, BiocStyle, RUnit, BiocGenerics
License:	Artistic-2.0
biocViews:	NetworkInference, NetworkEnrichment, GeneRegulation, GeneExpression, GraphAndNetwork

Index

MBR-class:	an S4 class for co-regulation analysis and inference of 'dual regulons'.
mbrPreprocess:	a preprocessing function for objects of class MBR.
mbrPermutation:	inference of transcriptional networks.
mbrBootstrap:	inference of consensus transcriptional networks.
mbrDpiFilter:	a filter based on the Data Processing Inequality (DPI) algorithm.
mbrAssociation:	motifs analysis and inference of "dual regulons".
mbrDuals:	a summary for results from the MBR methods.
tni2mbrPreprocess:	a preprocessing function for objects of class MBR.

Further information is available in the vignettes by typing `vignette("RTNduals")`. Documented topics are also available in HTML by typing `help.start()` and selecting the RTNduals package from the menu.

Author(s)

Vinicius S. Chagas, Clarice S. Groeneveld, Kerstin B Meyer, Gordon Robertson, Mauro A. A. Castro

References

Fletcher M.N.C. et al., *Master regulators of FGFR2 signalling and breast cancer risk.* Nature Communications, 4:2464, 2013.

Castro M.A.A. et al., *Regulators of genetic risk of breast cancer identified by integrative network analysis.* Nature Genetics, 48:12-21, 2016.

`MBR-class` *MBR objects*

Description

MBR: an S4 class for co-regulation analysis and inference of 'dual regulons'.

Details

The MBR class is a container for results from the MBR methods. The class slots are used to store information of different transcriptional networks, regulator annotation, infered 'dual regulons' and parameters used in the analysis. All the information is stored in nine slots.

Slots

`TNI1` a 'TNI' object created by the RTN package.

`TNI2` another 'TNI' object created by the RTN package.

`testedElementsTNI1` regulatory elements listed in the TNI1.

`testedElementsTNI2` regulatory elements listed in the TNI2.

`dualRegulons` all possible 'duals regulons' computed by `mbrAssociation`

`results` a list, results from the MBR methods.

`para` a list, parameters used in the MBR methods.

`summary` a list, summary for 'para' and 'results'.

`status` a character vector specifying the status of the MBR object based on the available methods.

Constructor

There are two constructors to create an MBR object, users can opt for one of the following: (1) `mbrPreprocess`; (2) `tni2mbrPreprocess`.

- (1): It is used to create an MBR object without any pre-computed transcriptional network.
- (2): It is used to create an MBR object using available transcriptional networks.

mbrAssociation,MBR-method

Motifs analysis and inference of 'dual regulons'.

Description

This function takes an MBR object and compares the shared regulon targets in order to test whether regulon pairs agree on the predicted downstream effects.

Usage

```
## S4 method for signature 'MBR'
mbrAssociation(object, regulatoryElements1 = NULL,
  regulatoryElements2 = NULL, minRegulonSize = 30, prob = 0.95,
  estimator = "spearman", pAdjustMethod = "BH", verbose = TRUE)
```

Arguments

object	A processed object of class MBR evaluated by the methods mbrPermutation, mbrBootstrap and mbrDpiFilter.
regulatoryElements1	An optional character vector specifying which 'TNI1' regulatory elements should be evaluated. If 'NULL' all regulatory elements will be evaluated.
regulatoryElements2	An optional character vector specifying which 'TNI2' regulatory elements should be evaluated. If 'NULL' all regulatory elements will be evaluated.
minRegulonSize	A single integer or numeric value specifying the minimum number of elements in a regulon. Gene sets with fewer than this number are removed from the analysis.
prob	A numeric value, representing a quantile threshold applyed to the association metric used to infer 'dual regulons'.
estimator	A character value specifying the estimator used in the association analysis. One of "spearman" (default), "kendall", or "pearson".
pAdjustMethod	A single character value specifying the p-value adjustment method to be used (see 'p.adjust' function for details).
verbose	A single logical value specifying to display detailed messages (when verbose=TRUE) or not (when verbose=FALSE).

Value

An MBR object with two data.frames in the slot 'results' listing the inferred 'dual regulons' and correspoding statistics.

Examples

```
##--- load a dataset for demonstration
data("dt4rtn", package = "RTN")
gexp <- dt4rtn$gexp
annot <- dt4rtn$gexpIDs
tfs1 <- dt4rtn$tfs[c("IRF8","IRF1","PRDM1","AFF3","E2F3")]
```

```
tfs2 <- dt4rtn$tfs[c("HCLS1","STAT4","STAT1","LMO4","ZNF552")]

##--- run mbrPreprocess
rmbr <- mbrPreprocess(gexp=gexp, regulatoryElements1 = tfs1,
regulatoryElements2=tfs2, gexpIDs=annot)

##--- run mbrPermutation
rmbr <- mbrPermutation(rmbr, nPermutations=10)

##--- run mbrBootstrap
rmbr <- mbrBootstrap(rmbr, nBootstrap=10)

##--- run mbrAssociation
rmbr <- mbrAssociation(rmbr, prob=0.75)
```

`mbrBootstrap,MBR-method`

Inference of consensus transcriptional networks.

Description

This function takes an MBR object and computes two consensus transcriptional networks.

Usage

```
## S4 method for signature 'MBR'
mbrBootstrap(object, verbose = TRUE, ...)
```

Arguments

`object`	A processed objec of class MBR evaluated by the method `mbrPermutation`.
`verbose`	A single logical value specifying to display detailed messages (when verbose=TRUE) or not (when verbose=FALSE).
`...`	Additional arguments passed to the `tni.bootstrap` function.

Value

An MBR object with two consensus mutual information matrices, one in each "TNI" slot.

Examples

```
##--- load a dataset for demonstration
data("dt4rtn", package = "RTN")
gexp <- dt4rtn$gexp
annot <- dt4rtn$gexpIDs
tfs1 <- dt4rtn$tfs[c("IRF8","IRF1","PRDM1","AFF3","E2F3")]
tfs2 <- dt4rtn$tfs[c("HCLS1","STAT4","STAT1","LMO4","ZNF552")]

##--- run mbrPreprocess
rmbr <- mbrPreprocess(gexp=gexp, regulatoryElements1 = tfs1,
regulatoryElements2=tfs2, gexpIDs=annot)
```

```
##--- run mbrPermutation
rmbr <- mbrPermutation(rmbr, nPermutations=10)

##--- run mbrBootstrap
rmbr <- mbrBootstrap(rmbr, nBootstrap=10)
```

---

mbrDpiFilter,MBR-method

A filter based on the Data Processing Inequality (DPI) algorithm.

Description

This function takes an MBR object and computes two transcriptional networks filtered by the data processing inequality algorithm.

Usage

```
## S4 method for signature 'MBR'
mbrDpiFilter(object, verbose = TRUE, ...)
```

Arguments

object	A processed object of class MBR evaluated by the methods mbrPermutation and mbrBootstrap.
verbose	A single logical value specifying to display detailed messages (when verbose=TRUE) or not (when verbose=FALSE).
...	Additional arguments passed to the tni.dpi.filter function.

Value

An MBR object with two DPI-filtered mutual information matrices, one in each "TNI" slot.

Examples

```
##--- load a dataset for demonstration
data("dt4rtn", package = "RTN")
gexp <- dt4rtn$gexp
annot <- dt4rtn$gexpIDs
tfs1 <- dt4rtn$tfs[c("IRF8","IRF1","PRDM1","AFF3","E2F3")]
tfs2 <- dt4rtn$tfs[c("HCLS1","STAT4","STAT1","LMO4","ZNF552")]

##--- run mbrPreprocess
rmbr <- mbrPreprocess(gexp=gexp, regulatoryElements1 = tfs1,
regulatoryElements2=tfs2, gexpIDs=annot)

##--- run mbrPermutation
rmbr <- mbrPermutation(rmbr, nPermutations=10)

##--- run mbrBootstrap
rmbr <- mbrBootstrap(rmbr, nBootstrap=10)

##--- run mbrDpiFilter
```

```
rmbr <- mbrDpiFilter(rmbr)
```

`mbrDuals,MBR-method` *A summary for results from the MBR methods.*

Description

This function lists the inferred 'dual regulons' and, if available, adds external evidences.

Usage

```
## S4 method for signature 'MBR'
mbrDuals(object, supplementary.table = NULL,
  evidenceColname = NULL, verbose = TRUE)
```

Arguments

`object`	A processed object of class MBR evaluated by the method `mbrAssociation`.
`supplementary.table`	An optional 'data.frame' with three columns representing (1) regulatory elements of 'TNI1', (2) regulatory elements of 'TNI2', and (3) external evidences between the regulatory elements.
`evidenceColname`	A single character value specifying a column in the 'supplementary.table'.
`verbose`	A single logical value specifying to display detailed messages (when verbose=TRUE) or not (when verbose=FALSE).

Value

An MBR object with an updated 'data.frame' in the slot 'results' listing the input additional evidences.

Examples

```
##--- load a dataset for demonstration
data("dt4rtn", package = "RTN")
gexp <- dt4rtn$gexp
annot <- dt4rtn$gexpIDs
tfs1 <- dt4rtn$tfs[c("IRF8","IRF1","PRDM1","AFF3","E2F3")]
tfs2 <- dt4rtn$tfs[c("HCLS1","STAT4","STAT1","LMO4","ZNF552")]

##--- run mbrAssociation
rmbr <- mbrPreprocess(gexp=gexp, regulatoryElements1 = tfs1,
regulatoryElements2=tfs2, gexpIDs=annot)
rmbr <- mbrPermutation(rmbr, nPermutations=10)
rmbr <- mbrBootstrap(rmbr, nBootstrap=10)
rmbr <- mbrAssociation(rmbr, prob=0.75)
rmbr <- mbrDuals(rmbr)

##--- check results
results <- mbrGet(rmbr, what="dualsInformation")
```

```
##--- add supplementary evidences
## here we build a 'toy' example using the 'rnorm' function for
## demonstration purposes only!
supplementaryTable <- results[,c("Regulon1","Regulon2")]
supplementaryTable$ToyEvidence <- rnorm(nrow(results))
supplementaryTable

##--- add supplementary evidences with brDuals function
rmbr <- mbrDuals(rmbr, supplementary.table=supplementaryTable,
evidenceColname = "ToyEvidence")

##--- check updated results
mbrGet(rmbr, what="dualsInformation")
```

`mbrGet,MBR-method` *Get information from individual slots in MBR object.*

Description

Get information from individual slots in an MBR object and any available results from previous analysis.

Usage

```
## S4 method for signature 'MBR'
mbrGet(object, what = "status")
```

Arguments

`object`	A preprocessed object of class MBR
`what`	a single character value specifying which information should be retrieved from the slots. Options: "TNI1", "TNI2", "testedElementsTNI1", "testedElementsTNI2", "dualRegulons", "results", "para", "summary", "status", "dualsInformation" and "hyperResults"

Value

A slot content from a object of class 'MBR' MBR object

Examples

```
##--- load a dataset for demonstration
data("dt4rtn", package = "RTN")
gexp <- dt4rtn$gexp
annot <- dt4rtn$gexpIDs
tfs1 <- dt4rtn$tfs[c("IRF8","IRF1","PRDM1","AFF3","E2F3")]
tfs2 <- dt4rtn$tfs[c("HCLS1","STAT4","STAT1","LMO4","ZNF552")]

##--- run mbrPreprocess
rmbr <- mbrPreprocess(gexp=gexp, regulatoryElements1 = tfs1,
regulatoryElements2=tfs2, gexpIDs=annot)
```

```
##--- get the 'TNI1' slot using 'mbrGet'
tni1 <- mbrGet(rmbr, what="TNI1")
```

`mbrPermutation,MBR-method`

Inference of transcriptional networks.

Description

This function takes an MBR object and computes two transcriptional networks inferred by mutual information (with multiple hypothesis testing corrections).

Usage

```
## S4 method for signature 'MBR'
mbrPermutation(object, verbose = TRUE, ...)
```

Arguments

`object`	A preprocessed object of class MBR.
`verbose`	A single logical value specifying to display detailed messages (when verbose=TRUE) or not (when verbose=FALSE).
`...`	Additional arguments passed on to the `tni.permutation` function.

Value

An MBR object with two mutual information matrices, one in each "TNI" slot.

Examples

```
##--- load a dataset for demonstration
data("dt4rtn", package = "RTN")
gexp <- dt4rtn$gexp
annot <- dt4rtn$gexpIDs
tfs1 <- dt4rtn$tfs[c("IRF8","IRF1","PRDM1","AFF3","E2F3")]
tfs2 <- dt4rtn$tfs[c("HCLS1","STAT4","STAT1","LMO4","ZNF552")]

##--- run mbrPreprocess
rmbr <- mbrPreprocess(gexp=gexp, regulatoryElements1 = tfs1,
regulatoryElements2=tfs2, gexpIDs=annot)

##--- run mbrPermutation
rmbr <- mbrPermutation(rmbr, nPermutations=10)
```

mbrPlotDuals *Plot shared target clouds between dual regulons.*

Description

This function plots the shared target clouds between a regulon pair.

Usage

```
mbrPlotDuals(object, names.duals = NULL, filepath = NULL, alpha = 0.8,
  lncols = c("darkgreen", "darkorange3"), lwd = 0.7, Pvalue = FALSE)
```

Arguments

object	A processed object of class MBR evaluated by the method mbrAssociation.
names.duals	A vector with 'dual regulon' indentifiers from the 'dualsInformation' table.
filepath	A character string indicating the file path where the plot should be saved.
alpha	The alpha transparency, a number in [0,1].
lncols	A vector of length 2 indicating the colors of the negative and positive target clouds, respectively.
lwd	Line width, a decimal value (between 0 and 1).
Pvalue	A Boolean value that indicates whether the 'dual regulon' p-value will be showed in the plot.

Value

A plot with the shared target clouds between dual regulons.

Examples

```
##--- load a dataset for demonstration
data("dt4rtn", package = "RTN")
gexp <- dt4rtn$gexp
annot <- dt4rtn$gexpIDs
tfs1 <- dt4rtn$tfs[c("IRF8","IRF1","PRDM1","AFF3","E2F3")]
tfs2 <- dt4rtn$tfs[c("HCLS1","STAT4","STAT1","LMO4","ZNF552")]

##--- run mbrPreprocess
rmbr <- mbrPreprocess(gexp=gexp, regulatoryElements1 = tfs1,
regulatoryElements2=tfs2, gexpIDs=annot)

##--- run mbrPermutation
rmbr <- mbrPermutation(rmbr, nPermutations=10)

##--- run mbrBootstrap
rmbr <- mbrBootstrap(rmbr, nBootstrap=10)

##--- run mbrAssociation
rmbr <- mbrAssociation(rmbr, prob=0.75)

##--- run mbrDuals
```

```
rmbr <- mbrDuals(rmbr)

##--- get inferred duals and plot the shared cloud of targets
duals <- mbrGet(rmbr, what="dualRegulons")
mbrPlotDuals(rmbr, names.duals=duals[1])
```

---

mbrPreprocess,matrix-method

A preprocessing function for objects of class MBR.

Description

A preprocessing function for objects of class MBR.

Usage

```
## S4 method for signature 'matrix'
mbrPreprocess(gexp, regulatoryElements1, regulatoryElements2,
  verbose = TRUE, ...)
```

Arguments

gexp	A numerical matrix, typically with mRNA and/or miRNA expression values.
regulatoryElements1	A named vector with regulatory elements listed in 'gexp' rownames.
regulatoryElements2	A named vector with regulatory elements listed in 'gexp' rownames.
verbose	A single logical value specifying to display detailed messages (when verbose=TRUE) or not (when verbose=FALSE).
...	Additional arguments passed on to tni.preprocess function.

Value

A preprocessed 'MBR-class' object.

Examples

```
##--- load a dataset for demonstration
data("dt4rtn", package = "RTN")
gexp <- dt4rtn$gexp
annot <- dt4rtn$gexpIDs
tfs1 <- dt4rtn$tfs[c("IRF8","IRF1","PRDM1","AFF3","E2F3")]
tfs2 <- dt4rtn$tfs[c("HCLS1","STAT4","STAT1","LMO4","ZNF552")]

##--- run mbrPreprocess
rmbr <- mbrPreprocess(gexp=gexp, regulatoryElements1 = tfs1,
regulatoryElements2=tfs2, gexpIDs=annot)
```

tni2mbrPreprocess,TNI-method *A preprocessing function for objects of class MBR.*

Description

This function merges two TNI class objects and creates one MBR class object.

Usage

```
## S4 method for signature 'TNI'
tni2mbrPreprocess(TNI1, TNI2, regulatoryElements1 = NULL,
  regulatoryElements2 = NULL, verbose = TRUE)
```

Arguments

TNI1	A 'TNI' class object.
TNI2	Another 'TNI' class object
regulatoryElements1	A character vector specifying which 'TNI1' regulatory elements should be evaluated.
regulatoryElements2	A character vector specifying which 'TNI2' regulatory elements should be evaluated.
verbose	A single logical value specifying to display detailed messages (when verbose=TRUE) or not (when verbose=FALSE).

Value

An MBR object.

Examples

```
#--- load a dataset for demonstration
data("dt4rtn", package = "RTN")
gexp <- dt4rtn$gexp
annot <- dt4rtn$gexpIDs
tfs1 <- dt4rtn$tfs[c("IRF8","IRF1","PRDM1","AFF3","E2F3")]
tfs2 <- dt4rtn$tfs[c("HCLS1","STAT4","STAT1","LMO4","ZNF552")]

## Not run:

##--- compute a TNI for tfs1
tni1 <- new("TNI", gexp=gexp, transcriptionFactors=tfs1)
tni1 <- tni.preprocess(tni1, gexpIDs=annot)
tni1 <-tni.permutation(tni1)
tni1 <-tni.bootstrap(tni1)

##--- compute a TNI for tfs2
tni2 <- new("TNI", gexp=gexp, transcriptionFactors=tfs2)
tni2 <- tni.preprocess(tni2, gexpIDs=annot)
tni2 <-tni.permutation(tni2)
```

```
tni2 <-tni.bootstrap(tni2)

##--- run tni2mbrPreprocess
rmbr <- tni2mbrPreprocess(tni1, tni2)

## End(Not run)
```

Index