aracne.networks, a data package containing gene regulatory networks assembled from TCGA data by the ARACNe algorithm

Federico M. Giorgi[1,2], Mariano J. Alvarez[1,3], and Andrea Califano[1]

[1]Department of Systems Biology, Columbia University, New York, USA
[2]CRUK, Cambridge University, Cambridge, UK
[3]DarwinHealth Inc., New York, USA

October 31, 2024

1 Overview of aracne.networks data package

The *aracne.networks* data package provides context-specific transcriptional regulatory networks (also called interactomes or regulons) reverse engineered by the ARACNe algorithm from The Cancer Genome Atlas (TCGA) RNAseq expression profiles.

ARACNe networks This package contains 25 Mutual Information-based networks assembled by ARACNe-AP [1] with default parameters (MI p-value = 10^{-8}, 100 bootstraps and permutation seed = 1). ARACNe is a network inference algorithm based on an Adaptive Partioning (AP) Mutual Information (MI) approach [1]. In short, ARACNe-AP estimates all pairwise Mutual Information scores between gene expression profiles, then assesses the significance of such Mutual Information by comparison to a null dataset. ARACNe then draws network edges between centroid genes (Transcription Factors and Signaling Proteins) and genes significantly associated with them (i.e. with significant MI). It then calculates Data Processing Inequality (DPI) to reduce the number of indirect connections.

ARACNe-AP was run on RNA-Seq datasets normalized using Variance-Stabilizing Transformation [2]. The raw data was downloaded on April 15$^{\text{th}}$, 2015 from the TCGA official website [3]. We follow the TCGA naming convention (e.g. BRCA = Breast Carcinoma) to name the individual context-specific networks.

```
> library(aracne.networks)
> data(package="aracne.networks")$results[, "Item"]
```

```
 [1] "regulonblca" "regulonbrca" "reguloncesc" "reguloncoad" "regulonesca"
 [6] "regulongbm"  "regulonhnsc" "regulonkirc" "regulonkirp" "regulonlaml"
[11] "regulonlihc" "regulonluad" "regulonlusc" "regulonnet"  "regulonov"
[16] "regulonpaad" "regulonpcpg" "regulonprad" "regulonread" "regulonsarc"
[21] "regulonstad" "regulontgct" "regulonthca" "regulonthym" "regulonucec"
```

Write a network to file The package contains a function to print individual networks into a file. Four columns will be printed: the Regulator id, the Target id, the Mode of Action (MoA, inferred by Spearman correlation analysis [4]) that indicates the sign of the association between regulator and target gene and ranges betrween -1 and +1, the Likelihood (essentially an edge weight that indicates how strong the mutual information for an edge is when compared to the maximum observed MI in the network, it ranges between 0 and 1). Further details about the *regulon* object as a model for transcriptional regulation are present in the manuscript [4].

In the following example, we print the first 10 interactions from the bladder carcinoma (blca) network. The network genes are identified by Entrez Gene ids.

```
> data(regulonblca)
> write.regulon(regulonblca, n = 10)
```

```
Regulator        Target        MoA        likelihood
10002        2648        0.994689591270463        0.886774633189913
10002        677827        0.116175345640136        0.707841406455471
10002        80152        0.999770437015603        0.950286744281199
10002        284382        -0.0368424333564396        0.0419762049859333
10002        9866        0.972066598154448        0.442238853411591
10002        283422        -0.574084929385018        0.260828476620346
10002        221613        -0.0959242601820319        0.717904706549976
10002        348174        0.953943934091558        0.814491117578869
10002        373509        0.704691385719852        0.244337186726846
10002        8803        -0.959165656086931        0.831653033754096
```

The user may want to analyze all the connections of a particular regulator (E.g. "399", the RHOH gene).

```
> data(regulonblca)
> write.regulon(regulonblca, regulator="399")
```

```
Regulator        Target        MoA        likelihood
399        9595        1        0.999999439751274
399        54440        1        0.999999439753891
399        5788        1        0.999993691255193
399        2124        1        0.999993972431349
399        10563        0.999999999999987        0.999880973084544
399        80342        1        0.999979237947268
399        1840        0.999999959099145        0.994240739975982
399        8875        0.999999999999397        0.999602389369848
399        6689        0.999999999998723        0.999531614767901
399        200186        0.154403590654008        0.948828817305409
399        165631        0.999999999950565        0.998777586463862
399        54509        0.999999981560018        0.997883918024065
399        171389        0.999999994824044        0.996800613785205
399        147929        -0.999154534552766        0.985197674740525
399        23416        0.999929331217517        0.96812145442081
399        26015        -0.992838466368412        0.834785111763068
399        10148        0.999999999999872        0.999729153685544
399        4951        -0.0504647730526015        0.544073601564966
399        57003        -0.0751708929022855        0.714920200879607
```

References

[1] Giorgi,F.M. et al. (2016) ARACNe-AP: Gene Network Reverse Engineering through Adaptive Partitioning inference of Mutual Information. Bioinformatics doi: 10.1093/bioinformatics/btw216.

[2] Anders, S and Huber W. (2010) Differential expression analysis for sequence count data. Genome Biol 2010;11(10):R106

[3] Weinstein J.N. et al. (2013) The cancer genome atlas pan-cancer analysis project. Nature Genetics 45, 1113-1120 2013

[4] Alvarez M.J. et al. (2016) Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nature Genetics *in press*.