Using DeMAND, a package for Interrogating Drug Mechanism of Action using Network Dysregulation analysis

Jung Hoon Woo, Yishai Shimoni, Mukesh Bansal, and Andrea Califano
Columbia University, New York, USA

October 29, 2019

1 Overview of DeMAND

Characterization of compound Mechanism of Action (MoA) is a critical, albeit complex and lengthy component of the drug development pipeline [1]. It is relevant both to determine the specificity of on-target activity as well as to identify off-target effects associated with potential toxicity, thus providing critical insight into the two major challenges of drug development. The MoA of a compound is defined as the set of potentially cell-context-specific biochemical interactors and effector through which a compound produces its pharmacological effects. Most of the experimental approaches relying on direct binding assays such as affinity purification [2, 3] or affinity chromatography assays [4]. These methods are limited to the identification of drug substrates with strong binding affinity to their ligand, rather than the full repertoire of proteins that effect compound activity in a specific tissue. Recently, systematic profiling of gene expression profiles (GEP) following compound perturbation in cell lines [5] has spurred development of novel computational methods for MoA analysis. The most advanced genomics based MoA studies have been conducted by network-based methods which have been recently proposed [6, 7]. However, these methods either rely on prior knowledge of the specific sub-networks that most likely mediate compound activity (i.e., not suitable for genome-wide analyses) or require a large number of samples (N > 100) to assess compound-specific rewiring of network topology thus limiting their practical utility. To address these challenges, we introduce a novel algorithm for Detecting Mechanism of Action based Network Dysregulation (DeMAND). The algorithm interrogates a pre-defined tissue-specific regulatory network, using a relatively small number of GEPs (N > 6) representing in vitro or in vivo compound-specific perturbations, to identify compound targets and effectors. The method is based on the realization that drugs affect the protein activity of their targets, but not necessarily their mRNA expression levels. In contrast, the change in protein activity directly affects the mRNA expression levels of downstream genes. Based on this hypothesis, DeMAND identifies drug MoA by comparing gene expression profiles following drug perturbation with control samples, and computing the change in the individual interactions within a pre-determined integrated transcriptional and post-translational regulatory model (interactome). For each edge in the interactome we determine the two-dimensional probability distribution of the gene expression levels both in the control state, and following drug treatment. Any changes in the probability distribution are estimated using the Kullback-Leibler (KL) divergence, from which we determine the statistical significance of the dysregulation of each edge. In the second step of DeMAND, we interrogate each gene independently to determine whether its interactions are enriched in dysregulated ones, suggesting that it is a candidate mechanism of action[8].

2 Citation

Woo JH, Shimoni Y, Yang WS, Subramaniam P, Iyer A, Nicoletti P, Martinez MR, Lopez G, Mattioli M, Realubit R, Karan C, Stockwell BR, Bansal M, Califano A (2015) Elucidating Compound Mechanism of Action by Network Dysregulation Analysis in Perturbed Cells (under review)

3 Installation of *DeMAND* package

In order to install the *DeMAND* package, the user needs to first install R (`http://www.r-project.org`). After that *DeMAND* and the required packages can be installed with:

```
> if (!requireNamespace("BiocManager", quietly=TRUE))
+     install.packages("BiocManager")
> BiocManager::install("DeMAND")
```

4 To run DeMAND using the data provided the *DeMAND* package

4.1 Getting Started

After installing the *DeMAND* package, DeMAND can be loaded by

```
> library(DeMAND)
```

4.2 Loading the data and generating a demand object

The data distributed in the *DeMAND* package is required to execute the code provided in this vignette. The data contains a gene expression dataset, annotation of the expression data, sample indices, and regulatory network:

bcellExp A subest of a gene expression profiles from DLBCL cells treated by Geldanamycin and by DMSO as control.

bcellAnno Annotation information for the probes of the gene expression matrix

bcellNetwork A subset of a molecular interaction network of Bcell assembled by the ARACNe[9] algorithm for protein-DN interactions and Bayesian method[10] for protein-protein interactions.

caseIndex Column indices of the gene expression matrix for the samples treated by Geldanamycin.

controlIndex Column indices of the gene expression matrix for the samples treated by DMSO.

The data provided in the *DeMAND* package can be loaded into memory with:

```
> data(inputExample)
> #ls()
```

Then, we can create an instance of class "demand" and store the required data with:

```
> dobj <- demandClass(exp=bcellExp, anno=bcellAnno, network=bcellNetwork)
> printDeMAND(dobj)
```

4.3 Drug MoA prediction using the *runDeMAND* function

We consider a gene as a part of Drug MoA if the interactions surrounding the gene are significanly dysregulated. The `runDeMAND()` function performs the prediction and store the results in the *demand* object.

```
> dobj <- runDeMAND(dobj, fgIndex=caseIndex, bgIndex=controlIndex)
> printDeMAND(dobj)
```

```
  moaGene        Pvalue           FDR
HSP90AB1 7.143445e-16 5.929059e-14
HSP90AA1 4.122803e-14 1.710963e-12
   PRKDC 3.709529e-03 9.646847e-02
    AKT1 7.206213e-03 9.646847e-02
   CREB1 8.056401e-03 9.646847e-02
   IKBKB 9.200592e-03 9.646847e-02
   gene1    gene2                KLD                   KLD.p
HSP90AB1      MDH1 19.4234259772248 0.00813939018047754
HSP90AB1       MYC 18.7243522480777 0.00881425862440678
HSP90AA1     HSPA8 18.3680838814313 0.00919001020472924
    ESR1 HSP90AB1 17.4293319382159  0.0102997424801221
HSP90AB1     PRKDC  15.910295299091  0.0125570388644811
HSP90AB1      TBX5 15.1001135358549  0.0140673635309947
```

Here are some information and the requirements for you to generate a *demand* object and to run the `runDeMAND()` function.

exp A N-by-M numeric matrix where the rows represent N probes (or genes) and the columns represent M samples.

anno A N-by-2 character matrix where the rows represent probes or genes in the same order as the *exp* matrix. The first column must hold the probe id or gene name as appears in the *exp* matrix, and the second column should hold their corresponding names (e.g gene symbol) as appears in the *network* matrix.

network A K-by-L (L>1) character matrix containing K interactions. The 1st column and the 2nd column contain the names of the interacting genes. The following columns may include additional data about the interaction, but are currently ignored.

fgIndex A numeric vector contains indices of columns which represent case samples (e.g. drug treated). The sample size should be greater than 3.

bgIndex A numeric vector contains indices of columns which represent control samples (e.g. drug treated). The sample size should be greater than 3.

5   References

References

[1] Scannell, J.W., et al., Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov, 2012. 11(3): p. 191-200.

[2] Ito, T., et al., Identification of a Primary Target of Thalidomide Teratogenicity. Science, 2010. 327(5971): p. 1345-1350

[3] Hirota, T., et al., Identification of Small Molecule Activators of Cryptochrome. Science, 2012. 337(6098): p. 1094-1097.

[4] Aebersold, R. and M. Mann, Mass spectrometry-based proteomics. Nature, 2003. 422(6928): p. 198-207.

[5] Lamb, J., et al., The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science, 2006. 313(5795): p. 1929-35.

[6] Gardner, T.S., et al., Inferring genetic networks and identifying compound mode of action via expression profiling. Science, 2003. 301(5629): p. 102-5.

[7] di Bernardo, D., et al., Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. Nat Biotechnol, 2005. 23(3): p. 377-83.

[8] Woo, J.H., et al., Elucidating Compound Mechanism of Action by Network Perturbation Analysis. Cell 2015. 162(2): p. 1-11.

[9] Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. 2006. BMC Bioinformatics 7 Suppl 1, S7.

[10] Lefebvre, C. et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol. Syst. Biol. 2010. 6, 377.

6   Session information

The output in this vignette was produced under the following conditions:

```
> sessionInfo()

R version 3.6.1 (2019-07-05)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 18.04.3 LTS

Matrix products: default
BLAS:   /home/biocbuild/bbs-3.10-bioc/R/lib/libRblas.so
LAPACK: /home/biocbuild/bbs-3.10-bioc/R/lib/libRlapack.so

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_US.UTF-8        LC_COLLATE=C
 [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] DeMAND_1.16.0      KernSmooth_2.23-16

loaded via a namespace (and not attached):
[1] compiler_3.6.1 tools_3.6.1
```