

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

January 16, 2015

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:

```
> source("http://bioconductor.org/biocLite.R")
> biocLite("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 subset 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)
```

An object of class demand

Slot exp:

Expression data of 1577 features by 12 samples

Slot anno:

Annotation of 1577 probes and 452 genes

Slot network:

Interactome with 2684 nodes, 1577 edges

Slot moa (head):

Empty

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 significantly dysregulated. The `runDeMAND()` function performs the prediction and store the results in the *demand* object.

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

An object of class demand

Slot exp:

Expression data of 1577 features by 12 samples

Slot anno:

Annotation of 1577 probes and 452 genes

Slot network:

Interactome with 2684 nodes, 1577 edges

Slot moa (head):

	moaGene	Pvalue	adjustedPvalue
18	HSP90AA1	1.071840e-164	8.896274e-163
26	HSP90AB1	3.359700e-27	1.394276e-25
5	CREB1	6.201270e-21	1.715685e-19
39	PRKDC	1.223552e-20	2.538870e-19
4	CDK4	4.176776e-20	6.933449e-19
7	E2F4	1.339405e-19	1.852844e-18

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 and the rows are N probes and the columns are M samples.

anno A N by 2 character matrix. The rows are probes but the order should be the same with the *demand* matrix. The first column can be anything (usually probe IDs) but the second column should includes Official Gene Symbol information for each probe.

network A K by 4 character matrix which contains K interactions. The 1st column and the 2nd column contain pairs of interacted genes. The 3rd and 4th columns indicate whether the interactions are protein-protein interaction (ppi) or protein-DNA (pdi) interaction. Column name should be as follows: `c("Gene1", "Gene2", "ppi", "pdi")`

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 Dysregulation Analysis in Perturbed Cells. 2015. (under review)
- [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.