DDN2.0: R and Python packages for differential dependency network analysis of biological systems

Bai Zhang[1,†], Yi Fu[1,†], Yingzhou Lu[1,†], Zhen Zhang[2], Robert Clarke[3], Jennifer E. Van Eyk[4], David M. Herrington[5], and Yue Wang[1,#]

[1]Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, VA 22203, USA; [2]Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, MD 21231, USA; [3]The Hormel Institute, University of Minnesota, Austin, MN 55912, USA; [4]Advanced Clinical Biosystems Research Institute, Cedars Sinai Medical Center, Los Angeles, CA 90048, USA; [5]Department of Internal Medicine, Wake Forest University, Winston-Salem, NC 27157, USA

[†] These authors are with equal contributions

[#] Author for correspondence: Yue Wang, Ph.D.
Virginia Tech Research Center - Arlington
900 N. Glebe Road, Arlington, VA 22203
E-mail: yuewang@vt.edu

Summary

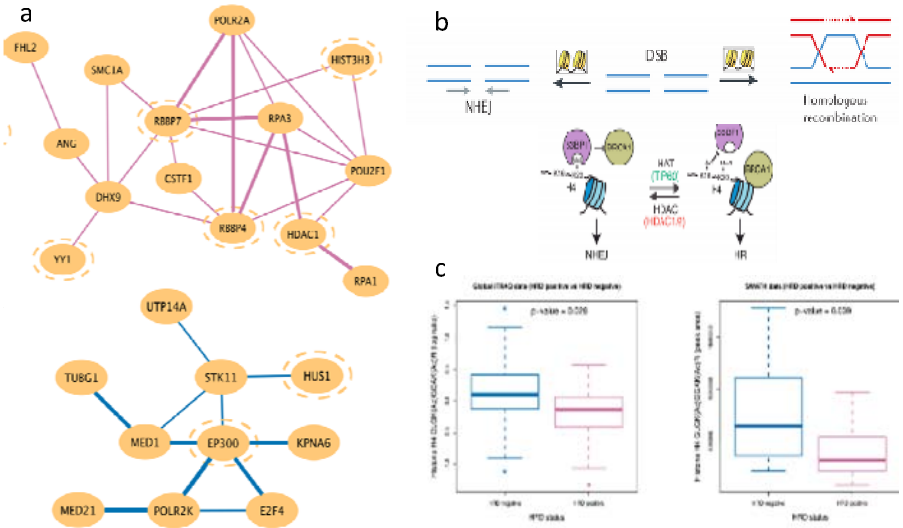
Data-driven differential dependency network analysis identifies in a complex and often unknown overall molecular circuitry a network of differentially connected molecular entities (pairwise selective coupling or uncoupling depending on the specific phenotypes or experimental conditions). Such differential dependency networks are typically used to assist in the inference of potential key pathways. Based on our previously developed Differential Dependency Network (DDN) method, we report here the fully implemented R and Python software tool packages for public use. The DDN algorithm uses a fused Lasso model and block-wise coordinate descent to estimate both the common and differential edges of dependency networks. The identified DDN can help to provide plausible interpretation of data, gain new insight of disease biology, and generate novel hypotheses for further validation and investigations.

Availability and Implementation: The open-source R and Python codes are freely available for download at https://github.com/MintaYLu/DDN

Contact: yuewang@vt.edu, lyz66@vt.edu

Supplementary information: Supplementary data are available upon request.

Data-driven differential network analysis identifies in a complex and often unknown overall molecular regulatory circuitry a network of differentially connected molecular entities (**Fig. 1a**; pairwise selective coupling or uncoupling depending on the specific phenotypes or experimental conditions, including the associated regulatory elements) (Ha, et al., 2015; Herrington, et al., 2018; Hu, et al., 2016; Mitra, et al., 2013; Ottenbros, et al., 2021; Tolios, et al., 2020). Such differential regulatory networks are typically used to assist in the inference of potential key pathways and targets in tumor biology. They can serve as useful frameworks for the construction and verification of mechanistic cancer models, which in turn help to provide plausible interpretation of data, gain new insight of cancer biology, and generate hypotheses for further validation and investigations (Califano, 2011; Gill, et al., 2010; Hudson, et al., 2009; Ideker and Krogan, 2012; Zhang, et al., 2016). It should be noted that the pairwise selective coupling or uncoupling in such data-derived network are based on statistical significance in differential co-expressions, representing possible underlying mechanisms involving direct or indirect, and single or collective effect of a multitude of regulatory pathways.



The publicly available Differential Dependency Network (DDN) analysis software tool (Zhang, et al., 2009; Zhang, et al., 2011; Zhang and Wang, 2010) and its recent extension, knowledge-fused DDN (kDDN) (Tian, et al., 2014), represent the effort to develop novel

bioinformatics tools to detect statistically significant and robust subnetworks with phenotype-specific changes. We have successfully applied DDN tools to real gene expression and proteomics data with promising findings and novel hypotheses (Crawford, et al., 2010; Hu, et al., 2015; Tian, et al., 2014; Zhang, et al., 2009; Zhang, et al., 2016). The current DDN tool can infer statistically significant coupling or uncoupling between pairs of molecular entities (e.g., genes/ proteins) that are dependent on the specific phenotypes or experimental conditions using transcriptome/proteome expression data (Tian, et al., 2014; Zhang, et al., 2009; Zhang, et al., 2011). To assess the DDN's effectiveness, we have compared the accuracy of inferring differential networks by DDN with other tools using realistic simulations (Tian, et al., 2014; Zhang, et al., 2009). The DDN tool has been used to detect critical changes in molecular networks in breast cancer (Madhavan, et al., 2011), ovarian cancer (Zhang, et al., 2016), and medulloblastoma . This has led to novel findings and hypotheses that have been validated experimentally (Hu, et al., 2015; Zhang, et al., 2016).

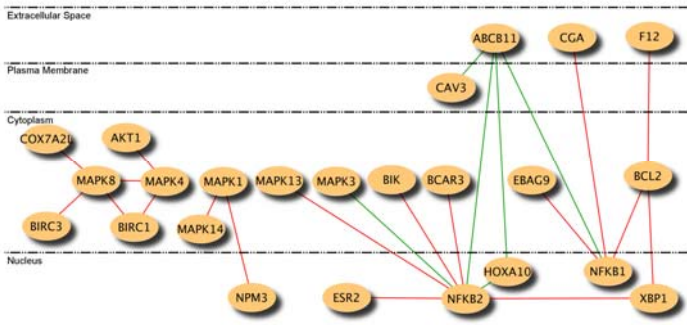
In our recent work on integrated proteogenomic characterization of human high grade serous ovarian cancer (Zhang, et al., 2016), applying DDN on a set of 171 BRCA1/2-related proteins, we identified a sub-network of 30 proteins that displayed differential co-expression patterns between HRD (homologous recombination deficiency) from non-HRD patients (**Fig. 1a**). Several of the proteins in these modules are known to be involved in histone acetylation or deacetylation. Although statistical association predicted by DDN cannot distinguish between drivers and consequences of HRD status, the observed enrichment of proteins associated with histone acetylation motivated us to identify and quantify acetylated peptides from our CPTAC global proteomic data. Comparative analysis of 399 acetylated peptides identified 15 acetylated peptides with significant differences between the HRD and non-HRD samples (dual acetylation at K12 and K16 of histone H4). Acetylation of H4 is involved in the choice of DNA double strand break (DSB) repair pathways (homologous recombination or non-homologous end joining). This relationship is regulated partially by HDAC1 (**Fig. 1b**). The findings were verified using independent samples and independent technical approaches (**Fig. 1c**) (Zhang, et al., 2016). The combined observations of increased HDAC1 (a hub protein also identified by DDN) and associated proteins at the pathway level, together with decreased acetylation of H4 in HRD patients at the (post-translational modifications) PTM level, provide insight regarding the potential role of HDAC1 in modulating the choice of DSB repair pathways (Zhang, et al., 2016).

From a scientific perspective, the main role of DDN is at least two twofold. First, an enrichment analysis focused on the genes/proteins that are differentially connected (nodes of differential networks including associated regulatory elements), not simply differentially expressed, can identify the more relevant pathways that functionally control phenotypes (Barabasi, et al., 2011; Califano, 2011; Hu, et al., 2015; Mitra, et al., 2013). Second, the hubs of differential networks (densely and differentially connected nodes) can hint at changes in the regulatory mechanisms, e.g., PTM or mutations in the regulatory elements of master regulators (Hudson, et al., 2009; Zhang, et al., 2016). By contrast, classic differential or co-expression analysis (i.e., identifying a subset of differentially or co- expressed genes/proteins) is insufficient here because the low level of transcription factors (TF) differential expression makes their detection challenging, even though TFs play a central regulatory role in controlling gene expression (Reverter, et al., 2010). DDN addresses an important biological issue because it better accounts for the functional activation/deactivation of TF (e.g., reversible phosphorylation, enhancer binding, chromatin openness) than does differential expression alone. Simple differential expression will overlook these vital changes in regulatory mechanism. For example, Hudson et al. reported that the myostatin gene containing a causal mutation was not detected because myostatin was not differentially expressed at any of ten developmental time points under surveillance (Hudson, et al., 2009).

The goal of studying network remodeling is not to solve the entire and common network at once but to find small, most-relevant, and critical features (e.g, topological changes within/between pathway networks) that can explain how the network functions to control phenotype. For example, most existing approaches infer gene co-expression networks under a single biological condition and are focused on the disease-related pathways enriched on a subset of differentially expressed genes. In marked contrast, DDN infers differential regulatory networks, i.e., differentially connected genes between different cancer subtypes, and is focused on the disease-related pathways enriched on (i) a subset of multi-omics expressions whose regulatory patterns (e.g., paired miRNA-mRNA/protein co-expressions) meet biological expectations and/or (ii) a subset of differentially connected genes/proteins (selectively coupled or uncoupled protein/gene pairs and associated regulatory elements depending on the specific cancer subtype). Another advantage is that we obtain small/sparse experimentally tractable models, i.e., rather than complex hairballs, DDN generally produces models that allow for the

design of feasible wet lab experiments that enable mechanistic validation (Hu, et al., 2015; Tyson, et al., 2011; Zhang, et al., 2016).

Case study to validate phenotype-specific differential networks of breast cancer dormancy and/or resistance. We have been previously funded to study how gene networks may adapt to affect the biology of breast cancers. Our CCSB/CSBC project is focused on understanding how estrogen receptor-positive (ER+) breast cancer cells adapt to the stress of endocrine-based therapies. Our central hypothesis invokes a gene network that coordinately regulates the functions of a cell and determines and executes the cell's fate decision. Using DDN tool we were able to identify three small topological features, one of which independently reflected much of our prior knowledge despite not explicitly incorporating this knowledge in learning the models (Zhang, et al., 2009). We followed the predictions of this differential topology and validated fundamentally new insights into molecular signaling, e.g., direct regulation of BCL2 by XBP1 and the requirement of NFκB for XBP1 signaling to regulate the pro-survival cell fate outcome in the cellular context of antiestrogen treatment and resistance (E2 versus E2+ICI) (**Fig. 2**) (Clarke, et al., 2011; Crawford, et al., 2010; Hu, et al., 2015; Tyson, et al., 2011). Specifically, DDN tool identified a new non-canonical UPR signaling route that was not previously known. Canonical signaling in the UPR would have predicted that changes in JNK regulated BCL2 but DDN predicted that XBP1 (not JNK) regulated BCL2 (Zhang, et al., 2009). We then found (i) response elements to XBP1 in the BCL2 promoter, (ii) that cells over expressing endogenous XBP1 also had higher BCL2 expression, (iii) that knockdown of XBP1 in these cells with RNAi reduced BCL2 expression, (iv) that constitutive overexpression of the XBP1 cDNA increased BCL2 expression and (v) that knockdown of BCL2 phenocopies knockdown of XBP1 with respect to reversing endocrine resistance (Crawford, et al., 2010; Hu, et al., 2015). We went on to further explore the DDN model and further confirmed the role of XBP1 regulation of NFkB (also not



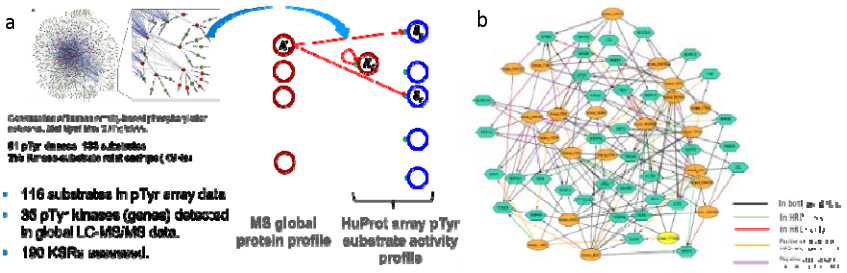
previously known) as another key driver of the prosurvival activities of XBP1 and the UPR (Hu, et al., 2015). In more detail, XBP1 can function as a direct/co-regulator through binding to its responsive element, most notably ERα. BCL2 contains response elements for both ERa and XBP1, thus XBP1-BCL2 may either be rewired or involve ERa as a latent variable, or intervening gene. In the non-recurring breast cancer, the affected network involves both signals received from activation of the membrane receptors and a cascade of signaling path inside a cell to promote apoptosis as well as survival. A balance between apoptosis and survival is necessary for damaged cells to be eliminated and repaired cells to survive. Three immune response genes (IL1B, NFκB and TNFα) for increased resistance to breast cancer treatment were identified in the recurring tumors. These genes formed a path to inhibit pro-apoptotic CASP3 and PPP3R1, and to activate pro-survival gene PIK3R5 or CSF2RB.

Case study to validation of DDN prediction on CPTAC data and identify phenotype-associated co-regulatory subnetworks. In our recent work (Zhang, et al., 2016), we reported a key finding on proteogenomic analysis of TCGA ovarian tumors in which DDN played a critical role (**Fig. 1**). The differentially networked proteins identified by DDN analysis on global proteomics data suggest that histone acetylation might play a role in DNA repair. We studied the downstream effects and identified several peptides of histone H4 showing differentially acetylated K12 and K16. We validated these discoveries using SWATH proteomics technology, in addition to a separate yet direct evidence demonstrated on a cell line model. In more detail, DDN analysis on ovarian cancer samples helped identifying a sub-network of 30 proteins involved in histone acetylation or deacetylation (**Fig. 1**). In cell line data, acetylation of H4 has previously been reported to be involved in the choice of DNA double strand break (DSB) repair pathways (homologous recombination or non-homologous end joining). This relationship is regulated partially by HDAC1, a protein also identified in the DDN analysis. We observed a significant enrichment of HDAC1 and its co-regulated proteins in tumors with HRD and low H4 acetylation relative to non-HRD tumors with high H4 acetylation (DDN analysis: permutation tests, p value <0.05; differentially acetylated peptides: t-tests, p value <0.05 with an estimated FDR $< 0.5\%$ by bootstrap/permutation tests). The identification of these acetylation events associated with HRD may provide an alternative biomarker of HRD for patient selection in future clinical trials of HDAC inhibitors alone or in combination with PARP inhibition. This may help to resolve the current discrepancy between the initial observation of limited single

agent activity of HDAC inhibitors in ovarian cancer and more recent findings of a >40% response rate when used in combination with cytotoxic chemotherapy in platinum-resistant patients (Zhang, et al., 2016).

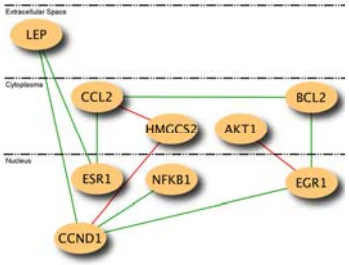
The integration of proteogenomic data also refers to the integration of molecular networks across different phenotypes/conditions to identify functional modules that are rewired in response to conditional fluxes. The types of networks themselves could be physical interaction networks, signaling and metabolic pathways, and others. The inference of conditional network often requires the ability to bridge between prior knowledge and experimental data and with built-in resampling/permutation to offer statistically sound and robust results (Zhang, et al., 2014). It should be noted that while DDN infers both common and differential networks, its unique and main function is to differentially detect the statistically significant "rewiring" associated with phenotypic differences. Thus, only rewired edges are included in the final differential dependency networks_that are not meant to represent the complete regulatory network that supports the entire functions of the phenotypes. By definition, the rewired part of the networks are not supposed to be huge. We emphasize that the entire molecular machinery behind HR repairing of DSB is much more completed than the DDN detected. However, this is the part that contains key elements differentiating HRD and non-HRD and pin-pointed to histone H4 peptide acetylation as a critical point of the mechanism.

Case study to analyze PTM interactome. The aforementioned findings of HRD status-dependent H4 acetylations, consistent with recent reports in the literature, indicate that much of the cancer associated molecular alterations will not necessarily be limited to the form of coordinated variations among DNA, copy number, and mRNA/protein expression, but rather as complex rewiring of signaling pathways/networks carried out in the protein and post-translational modification (PTM) interactomes. The DDN in its current form, and more so with



its proposed addition, can play an important role in differential analysis of PTM interactomes to capture phenotype/experiment condition-dependent changes. As an example, through CPTAC, we analyzed the tyrosine phosphorylation substrate activity data using the HuProt protein arrays(Hu, et al., 2014). The LC-MS/MS global protein profiles include those proteins that are known pTyr kinases. DDN was applied to identify subnetworks of highly correlated kinase (MS data) – substrate (HuProt array data) pairs that are dependent on HRD phenotypes. The identified subnetworks were further compared against the previously experimentally generated kinase-substrate relationship (KSR) map(Hu, et al., 2014) to retain only connections that are known possible KSRs (**Fig. 3a**). We were able to identify several kinase hubs, such as MAP2K4. Many of them are known drug targets (unpublished, on-going research) (**Fig. 3b**). In this example, the relationship between K and S nodes are directional (downstream substrates). The superposition of KSR map was applied separately. The extension of DDN to allow for seamless integration of such knowledge in a quantitative and computational way will further optimize such analysis approaches.

Case study on cancer prevention research. DDN tool has generated biologically meaningful hypotheses that guided further investigations. To study the consequence of possible transcriptional re-programming regulated by promoter methylation status (effectors ER, BCL2, LEP, and EGR1; AKT1 can regulate methylation patterns in some promoters, *e.g.*, AKT1-EGR1 edge), revealed by DDN, we found that *in utero* estrogen exposures induce a rewired network pattern in the mammary glands of rodent offspring that predicts for resistance to endocrine therapies. Subsequent studies have shown that tumors that arise in these mammary glands are less responsive to Tamoxifen (TAM) (Hilakivi-Clarke, et al., 2017). <u>This represents the first study to explain why many ER+ breast cancers fail to respond (or respond and later recur) with TAM treatment</u> (**Fig. 4**).



References

Barabasi, A.L., Gulbahce, N. and Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011;12(1):56-68.

Califano, A. Rewiring makes the difference. *Mol Syst Biol* 2011;7:463.

Clarke, R.*, et al.* Endoplasmic reticulum stress, the unfolded protein response, and gene network modeling in antiestrogen resistant breast cancer. *Horm Mol Biol Clin Investig* 2011;5(1):35-44.

Crawford, A.C.*, et al.* Co-inhibition of BCL-W and BCL2 restores antiestrogen sensitivity through BECN1 and promotes an autophagy-associated necrosis. *PLoS One* 2010;5(1):e8604.

Gill, R., Datta, S. and Datta, S. A statistical framework for differential network analysis from microarray data. *BMC Bioinformatics* 2010;11:95.

Ha, M.J., Baladandayuthapani, V. and Do, K.A. DINGO: differential network analysis in genomics. *Bioinformatics* 2015;31(21):3413-3420.

Herrington, D.M.*, et al.* Proteomic Architecture of Human Coronary and Aortic Atherosclerosis. *Circulation* 2018;137(25):2741-2756.

Hilakivi-Clarke, L.*, et al.* Effects of In Utero Exposure to Ethinyl Estradiol on Tamoxifen Resistance and Breast Cancer Recurrence in a Preclinical Model. *J Natl Cancer Inst* 2017;109(1).

Hu, J.*, et al.* Global analysis of phosphorylation networks in humans. *Biochimica et biophysica acta* 2014;1844(1 Pt B):224-231.

Hu, J.X., Thomas, C.E. and Brunak, S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* 2016;17(10):615-629.

Hu, R.*, et al.* NF-kappaB signaling is required for XBP1 (unspliced and spliced)-mediated effects on antiestrogen responsiveness and cell fate decisions in breast cancer. *Mol Cell Biol* 2015;35(2):379-390.

Hudson, N.J., Reverter, A. and Dalrymple, B.P. A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput Biol* 2009;5(5):e1000382.

Ideker, T. and Krogan, N.J. Differential network biology. *Mol Syst Biol* 2012;8.

Madhavan, S.*, et al.* G-DOC: a systems medicine platform for personalized oncology. *Neoplasia* 2011;13(9):771-783.

Mitra, K.*, et al.* Integrative approaches for finding modular structure in biological networks. *Nat Rev Genet* 2013;14(10):719-732.

Ottenbros, I.*, et al.* Network Analysis to Identify Communities Among Multiple Exposure Biomarkers Measured at Birth in Three Flemish General Population Samples. *Front Public Health* 2021;9:590038.

Reverter, A.*, et al.* Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* 2010;26(7):896-904.

Tian, Y.*, et al.* Knowledge-fused differential dependency network models for detecting significant rewiring in biological networks. *BMC Syst Biol* 2014;8(1):87.

Tolios, A.*, et al.* Computational approaches in cancer multidrug resistance research: Identification of potential biomarkers, drug targets and drug-target interactions. *Drug Resist Updat* 2020;48:100662.

Tyson, J.J.*, et al.* Dynamic modelling of oestrogen signalling and cell fate in breast cancer cells. *Nat Rev Cancer* 2011;11(7):523-532.

Zhang, B.*, et al.* Differential Dependency Network Analysis to Identify Condition-Specific Topological Changes in Biological Networks. *Bioinformatics* 2009;25(4):526-532.

Zhang, B.*, et al.* DDN: a caBIG(R) analytical tool for differential network analysis. *Bioinformatics* 2011;27(7):1036-1038.

Zhang, B., Tian, Y. and Zhang, Z. Network biology in medicine and beyond. *Circ Cardiovasc Genet* 2014;7(4):536-547.
Zhang, B. and Wang, Y. Learning structural changes of Gaussian graphical models in controlled experiments. In, *Uncertainty in Artificial Intelligence (UAI 2010)*. 2010.
Zhang, H.*, et al.* Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* 2016;166(3):755-765.