hARACNe: improving the accuracy of regulatory model reverse engineering via higher-order data processing inequality tests

In Sock Jang, Adam Margolin and Andrea Califano

1Sage Bionetworks, 1100 Fairview Avenue North, Seattle, WA 98109, USA
2Department of Systems Biology, Biochemistry and Molecular Biophysics, Biomedical Informatics, and Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY 10032, USA

A key goal of systems biology is to elucidate molecular mechanisms associated with physiologic and pathologic phenotypes based on the systematic and genome-wide understanding of cell context-specific molecular interaction models. To this end, reverse engineering approaches have been used to systematically dissect regulatory interactions in a specific tissue, based on the availability of large molecular profile datasets, thus improving our mechanistic understanding of complex diseases, such as cancer. In this paper, we introduce high-order Algorithm for the Reconstruction of Accurate Cellular Network (hARACNe), an extension of the ARACNe algorithm for the dissection of transcriptional regulatory networks. ARACNe uses the data processing inequality (DPI), from information theory, to detect and prune indirect interactions that are unlikely to be mediated by an actual physical interaction. Whereas ARACNe considers only first-order indirect interactions, i.e. those mediated by only one extra regulator, hARACNe considers a generalized form of indirect interactions via two, three or more other regulators. We show that use of higher-order DPI resulted in significantly improved performance, based on transcription factor (TF)-specific ChIP-chip data, as well as on gene expression profile following RNAi-mediated TF silencing.

1. Introduction

Cellular phenotypes are determined by a complex web of physical interactions between gene products [1]. Modelling these relationships helps to organize the list of parts encoded in the genome into functional genetic networks, a crucial step towards the understanding of mechanisms contributing to normal cell physiology as well as of their dysregulation in disease.

With the advent of high-throughput technologies, a large amount of molecular profile data have been generated from large numbers of samples associated with a variety of diseases. Through use of reverse engineering algorithms, these data have shown great promise in the dissection of transcriptional regulatory networks on a genomic scale [2–10]. While the caveats associated with the use of gene expression data for transcriptional network inference have been well documented and studied throughout the past decade [11], a preponderance of studies and high-impact discoveries in systems biology have established such approaches as widely accepted tools in a systems biologist’s arsenal [6,12–18]. Reverse engineering approaches have been developed using the mathematical frameworks established in disciplines such as Bayesian networks [4,5,19], dynamical systems [8] and information theory [2,6,7,11–13,16,17,20]. For instance, the relevance network approach [2,21], one of the earliest proposed, assumes that genes co-expressed above statistical significance are more likely to represent regulatory interactions. Unfortunately, owing to long chains of regulatory events,
a large number of gene pairs may be co-expressed without necessarily implicating a physical interaction. Thus, relevance networks, whether based on mutual information or Spearman correlation, typically generate a large number of false positives.

A survey of such approaches and their relative merits is reviewed in [11] and are outside the scope of this article. Rather, the current study is intended to build on and improve the body of research associated with the information theoretic algorithm ARACNe (Algorithm for the Reconstruction of Accurate Cellular Network), which has emerged as a widely referenced approach that has been experimentally validated in numerous applications, leading to key biological discoveries [6,12–14].

ARACNe was developed to maintain the simplicity of relevance networks, while using rigorous information theoretic principles to eliminate the vast majority of indirect interactions (i.e. false positives). By limiting interactions only to pairs where at least one of the genes is a transcription factor (TF), ARACNe also addressed the issue of interaction directionality, since a TF can regulate a non-TF but the opposite is not true. Thus, the only undirected interactions in ARACNe networks are those between two TFs, where directionality cannot be disambiguated.

To eliminate indirect interactions, ARACNe uses the data processing inequality (DPI) theorem, from information theory, stating that information transferred directly (i.e. through a physical interaction) is always larger than information transferred indirectly, i.e. via an intermediary. Here, information is formally computed by the mutual information, \( I(x;y) = S(x) + S(y) - S(x,y) \), where \( S(u) \) represents the entropy of the variable \( u \). We note that the DPI does not apply for other measures, such as Spearman correlation, unless interactions are linear, which is clearly not the case in biological systems. Experimental validation shows that the DPI removes a vast majority of false positive interactions, leading to highly accurate regulatory models [6,7,20,22].

Given its simplicity and the fact that it could consider each interaction independently, ARACNe was the first reverse engineering algorithm to successfully scale up to the complex regulatory networks of mammalian cells. For instance, 90 per cent of ARACNe-inferred and experimentally tested MYC targets in human B cells were validated as directly regulated by MYC [6], validation of targets in high-grade human glioma confirmed 40 out of 50 tested interactions for the transcription factors C/EBP\( \beta \), Stat3, RUNX1 and BHLHB2 [12], and the algorithm could correctly dissect a large number of transcriptional interactions involved in the synergistic control of germinal centre B cell proliferative programmes by FOXM1 and MYB [14,23,24]. More importantly, interrogation of ARACNe-inferred regulatory networks has allowed elucidation of key drivers of normal and disease-related phenotypes [6,12,13,15].

In this study, we explore improvements based on the nested iterative application of higher-order DPI tests [25,26]. The resulting algorithm, hARACNe (higher-order ARACNe), is thus designed to identify and remove a significant number of false positive interactions that could not be identified by first-order DPI (DPI\(^1\)) analysis. Use of higher-order DPI analysis does not affect the generality of the method and is applicable to any network analysis, including those using static, time-course and even post-translationally predicted data [7,27].

Our analysis shows that hARACNe can systematically eliminate false positive interactions that were missed by DPI\(^1\) logic of ARACNe, thus significantly improving inferred TF–target interaction accuracy, based on MYC-binding data from ChIP-chip data [6,28] as well as gene expression profile (GEP) analysis following RNAi-mediated silencing of BCL6 [16]. Thus, our data show that hARACNe constitutes an advance in the identification of bona fide TF–target interactions of biological relevance.

2. Material and methods

2.1. Data sources

For this study, 254 previously published [6,17,20] GEPs were used, representing 17 distinct normal and tumour-related B-cell phenotypes from primary patient biopsies and tumour-derived cell lines. In addition, we used a set of 226 previously published [16] GEPs from human B-cell lymphoma, including normal samples from naive, memory and germinal centre B cells isolated from human tonsils and patient derived tumour samples, including diffused large B-cell lymphoma (DLBCL), follicular lymphoma and chronic lymphocytic leukaemia (CLL). The first dataset was profiled using the HG-U95Av2 GeneChip platform, whereas the second set was profiled using the HG-U133 Plus2.0 GeneChip (Affymetrix) [6,7,14,16,17,20]. All profiles discussed in this research are accessible from the Gene Expression Omnibus (GEO; National Center for Biotechnology Information), through GEO series accession nos. GSE2350 and GSE12195. The list of phenotypes is found in table 1.

Ramos and Mutu (human Burkitt's lymphoma) cell lines were analysed in ChIP-chip assays to identify the genes whose proximal promoter is bound the MYC protein. ChIP-chip significance analysis (CSA) was applied, where \( p \)-values were first derived for each probe from three replicate experiments, as described in [29]. Values of \( p \) were then integrated across a 500-base region surrounding the transcription start site of the gene, using a gamma cumulative distribution function. Each promoter was associated with the highest 500-base MYC-localization segment, and the false discovery rate (FDR) was computed using the Benjamini Hochberg procedure, as a function of gene rank. More precise procedures were described in [28].

In addition, DLBCL cell lines, including LY7, Pfeiffer and VAL, were profiled following lentivirus-mediated shRNA silencing of the BCL6 TF. The experimental procedures and conditions used to perform these experiments are described in [16]. Following BCL6 silencing, differentially expressed genes were identified (FDR < 0.05). ARACNe- and hARACNe-inferred BCL6 targets (BCL6 regulators) were then compared with differentially expressed genes.

For both the ChIP-chip and shRNA experiments, the procedures described earlier were used to infer a list of ‘positive’ TF–target interactions, which were compared with the predictions made by hARACNe and ARACNe to compute the number of true positives versus false positives removed by the additional pruning steps:

\[
\text{TP}_{\text{gain}} = \text{TP}(\text{hARACNe}) - \text{TP}(\text{ARACNe})
\]

\[
\text{FP}_{\text{gain}} = \text{FP}(\text{ARACNe}) - \text{FP}(\text{hARACNe}).
\]

For comparison with ChIP-chip experiments, TP and FP are the number of ARACNe- or hARACNe-inferred MYC targets that were identified as bound and not bound by MYC, respectively. For comparison with shRNA experiments, TP and FP are the number of ARACNe- or hARACNe-inferred BCL6 targets that were identified as differentially regulated or not differentially regulated, respectively, following BCL6 inhibition. We note that the decrease in true positives due to removing interactions is equivalent to the increase in false negatives.

As with most ‘gold standards’ used in assessment of systems biology applications, we note that both the ChIP-chip and
Table 1. Human B-cell phenotypes in the two GEPs.

<table>
<thead>
<tr>
<th>HG-U95 Av2</th>
<th>HG-U133 Plus2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cell type</strong></td>
<td><strong>no.</strong></td>
</tr>
<tr>
<td>diffuse large B-cell lymphoma</td>
<td>68</td>
</tr>
<tr>
<td>Burkitt’s lymphoma</td>
<td>33</td>
</tr>
<tr>
<td>follicular lymphoma</td>
<td>14</td>
</tr>
<tr>
<td>mantle cell lymphoma</td>
<td>8</td>
</tr>
<tr>
<td>B-cell chronic lymphocytic leukaemia</td>
<td>34</td>
</tr>
<tr>
<td>hairy cell leukaemia</td>
<td>16</td>
</tr>
<tr>
<td>multiple myeloma</td>
<td>4</td>
</tr>
<tr>
<td>Hodgkin’s lymphoma</td>
<td>4</td>
</tr>
<tr>
<td>primary effusion lymphoma</td>
<td>9</td>
</tr>
<tr>
<td>splenic lymphoma with villous lymphocytes</td>
<td>12</td>
</tr>
<tr>
<td>large cell lymphomas</td>
<td>5</td>
</tr>
<tr>
<td>Burkitt’s lymphoma type III</td>
<td>3</td>
</tr>
<tr>
<td>undefined</td>
<td>2</td>
</tr>
<tr>
<td>germinatal centre B cell</td>
<td>17</td>
</tr>
<tr>
<td>naive B cell</td>
<td>5</td>
</tr>
<tr>
<td>memory B cells</td>
<td>5</td>
</tr>
<tr>
<td>cord blood</td>
<td>5</td>
</tr>
<tr>
<td>total</td>
<td>254</td>
</tr>
</tbody>
</table>

shRNA experiments are imperfect tests to identify true TF–target regulatory interactions. In particular, each assay only measures one (necessary but not sufficient) criterion of a regulatory interaction, and therefore may over-estimate the number of interactions. Moreover, each assay is performed in only a subset of tumour sub-types that were profiled in the microarray dataset, and therefore may underestimate the number of interactions, as only those active in the tested sub-types will be detected. Thus, although the exact number of true and false positives reported for each method are inexact, the TFgain and FPgain statistics should represent fair tests of the relative performance of each method. Moreover, the ChIP-chip and shRNA experiments measure different aspects of TF–target interactions and were performed on different TFs; therefore, consistent performance increases in comparison with these two distinct experiments lend strong support to the benefit of a method.

2.2 hARACNe algorithm design

The DPI is a simple but powerful theorem that starts from the simple axiom that as information is transferred through a lossy network, it can only be reduced and never increased. A Markov chain is at the heart of the DPI concept. A Markov chain is a linear sequence of states such that knowledge of the state at position \( i_0 \) makes all states at position \( i > i_0 \) independent of states at position \( i < i_0 \). In other words, any state provides all the necessary information to infer downstream states in the chain \([26,30]\). In a gene regulatory network, a Markov chain may represent a sequence of regulatory interactions \( R_1 \rightarrow R_2 \rightarrow \cdots \rightarrow R_N \). In this case, DPI can remove indirect interactions of the type \( R_1 \rightarrow R_2 \rightarrow R_3 \). In the case of DPI, consistent performance increases in comparison with these two distinct experiments lend strong support to the benefit of a method.

2.2 hARACNe algorithm design

The DPI is a simple but powerful theorem that starts from the simple axiom that as information is transferred through a lossy network, it can only be reduced and never increased. A Markov chain is at the heart of the DPI concept. A Markov chain is a linear sequence of states such that knowledge of the state at position \( i_0 \) makes all states at position \( i > i_0 \) independent of states at position \( i < i_0 \). In other words, any state provides all the necessary information to infer downstream states in the chain \([26,30]\). In a gene regulatory network, a Markov chain may represent a sequence of regulatory interactions \( R_1 \rightarrow R_2 \rightarrow \cdots \rightarrow R_N \). In this case, DPI analysis can remove indirect interactions of the type \( R_1 \rightarrow R_2 \rightarrow R_3 \). In this case, DPI states that, if any information is lost through interactions (an obvious true statement for any biological regulatory cascade), then \( I(\R_1; \R_2; \R_3) \) is strictly smaller than both \( I(\R_1; \R_3) \) and \( I(\R_1; \R_2) \), i.e.

\[ I(\R_1; \R_2; \R_3) < \min(\R_1; \R_3, \R_1; \R_2), \quad \text{i.e.} \]

\[ I(\R_1; \R_2; \R_3) < \min(I(\R_1; \R_3), I(\R_1; \R_2)) = \text{DPI}(R_1 \Rightarrow R_2). \quad (2.1) \]

This means that the \( R_1 \rightarrow R_2 \rightarrow R_3 \) interaction is removed because it is indirect through \( R_1 \).

To extend DPI to higher-order DPIs, consider a Markovian quadruplet \( R_1 \rightarrow R_2 \rightarrow R_3 \rightarrow R_4 \). Then, the second-order DPI (DPI²) could be expressed as follows:

\[ I(\R_1; \R_2; \R_3) \leq \min(I(\R_1; \R_3), I(\R_1; \R_2), I(\R_1; \R_4), I(\R_2; \R_3)) \quad \text{with DPI}(R_1 \Rightarrow R_2) \text{ and DPI}(R_1 \Rightarrow R_3) \]

\[ = \text{DPI}(R_1; R_2, R_3, R_4). \quad (2.2) \]

Or, in other words, the \( R_1 \rightarrow R_3 \) interaction is removed as indirect through \( R_1, R_2 \).

Proof: It follows from the Markov property satisfying conditional independency among a Markov triplet as \( I(R_1; R_3 | R_2, R_1) = I(R_1; R_3 | R_2) = I(R_1; R_3 | R_2) = 0 \). Therefore, by repeatedly using the chain rule, \( I(R_1 R_3 R_2) = I(R_1 R_3 R_2, R_1) \). We obtain this inequality as follows:

\[ I(\R_1; \R_3) \leq I(\R_1; \R_3 | \R_2) \]

\[ = I(\R_1; \R_3 | \R_2) + I(\R_2; \R_1 | \R_3) = I(\R_1; \R_3 | \R_2) \]

\[ \leq I(\R_1; \R_3 | \R_2, \R_1) \]

\[ = I(\R_1; \R_3 | \R_2) + I(\R_1; \R_3 | \R_2) = I(\R_1; \R_3 | \R_2). \]

Figure 1b shows graphically how the DPI² can be nested iteratively to identify additional false positive interactions within a Markov quadruplet.
of k DPI3 and DPI4. Detect indirect interaction in quintuplet, and (f) where interactions from lower DPI to higher DPI. Thus, indirect interactions.

(DPI as follows:

1 1 1 1 1

þ

min (1(R1)R2) . . . (1(Rn)Rn/n+1))

with {1st DPI}. {2nd DPI}. . . . {n-1}th DPI

= DPI(R1 . . . Rn => Rn/n+1),

(2.3)

where {kth DPI} means the super set of all possible combinations of kth-order DPI. Figure 1c,d, for instance, shows examples for DPI2 and DPI3.

Higher-order DPI pruning process is always sequentially conducted after lower-order DPI is processed in advance, and pairwise mutual information between TFs and TGs are once computed conducted after lower-order DPI is processed in advance, and pairwise mutual information between TFs and TGs are once computed. We can apply the nth-order DPI (DPIn) with conditional independence among any sub-chain as follows:

I(R1; Rn+1) ≤ min((I(R1; R2)) . . . (I(Rn; Rn+1))),

with {1st DPI}, {2nd DPI}, . . . . {n-1}th DPI

where {nth DPI} means the super set of all possible combinations of nth-order DPI. Figure 1c,d, for instance, shows examples for DPI2 and DPI3. Higher-order DPI pruning process is always sequentially conducted after lower-order DPI is processed in advance, and pairwise mutual information between TFs and TGs are once computed conducted after lower-order DPI is processed in advance, and pairwise mutual information between TFs and TGs are once computed conducted after lower-order DPI is processed in advance, and pairwise mutual information between TFs and TGs are once computed.

3. Results

3.1. Reverse engineered transcriptional networks comparing ARACNe and hARACNe

The number of transcriptional interactions in the reconstructed networks was compared, and the distribution of TF targets is shown in Table 2. In the human B-cell data profiled with the Affymetrix U95 platform, there were 12 600 probe IDs, of which 1225 corresponded to TFs, representing 848 unique genes. ARACNe identified 155 526 transcriptional interactions at the probe ID level. After applying DPI2, 134 452 interactions were left and 21 074 were removed as indirect interactions. When DPI3 was applied, 132 018 interactions were left and an additional 2434 were removed as indirect. Analysis with fourth- and fifth-order DPI removed only 280 and no interaction, respectively. Thus, it appears that high-order DPI reaches saturation rapidly and that only DPI3 and DPI4 provide substantial false positive filtering power. Yet, hARACNe, DPI2 and above, have relatively high computational cost compared with efficiency of filtering power. With the second B-cell dataset, profiled using the Affymetrix U133 platform, there were 14 090 probe IDs, of which 1290 were TFs representing 1209 unique genes.

ARACNe identified 198 766 transcriptional interactions at the probe ID level. DPI2 produced 142 037 interactions, with 848 unique genes. ARACNe identified 155 526 transcriptional interactions at the probe ID level. After applying DPI2, 134 452 interactions were left and 21 074 were removed as indirect interactions. When DPI3 was applied, 132 018 interactions were left and an additional 2434 were removed as indirect. Analysis with fourth- and fifth-order DPI removed only 280 and no interaction, respectively. Thus, it appears that high-order DPI reaches saturation rapidly and that only DPI3 and DPI4 provide substantial false positive filtering power. Yet, hARACNe, DPI2 and above, have relatively high computational cost compared with efficiency of filtering power. With the second B-cell dataset, profiled using the Affymetrix U133 platform, there were 14 090 probe IDs, of which 1290 were TFs representing 1209 unique genes.

ARACNe identified 198 766 transcriptional interactions at the probe ID level. DPI2 produced 142 037 interactions, with 848 unique genes. ARACNe identified 155 526 transcriptional interactions at the probe ID level. After applying DPI2, 134 452 interactions were left and 21 074 were removed as indirect interactions. When DPI3 was applied, 132 018 interactions were left and an additional 2434 were removed as indirect. Analysis with fourth- and fifth-order DPI removed only 280 and no interaction, respectively. Thus, it appears that high-order DPI reaches saturation rapidly and that only DPI3 and DPI4 provide substantial false positive filtering power. Yet, hARACNe, DPI2 and above, have relatively high computational cost compared with efficiency of filtering power. With the second B-cell dataset, profiled using the Affymetrix U133 platform, there were 14 090 probe IDs, of which 1290 were TFs representing 1209 unique genes.

848 unique genes. ARACNe identified 155 526 transcriptional interactions at the probe ID level. After applying DPI2, 134 452 interactions were left and 21 074 were removed as indirect interactions. When DPI3 was applied, 132 018 interactions were left and an additional 2434 were removed as indirect. Analysis with fourth- and fifth-order DPI removed only 280 and no interaction, respectively. Thus, it appears that high-order DPI reaches saturation rapidly and that only DPI3 and DPI4 provide substantial false positive filtering power. Yet, hARACNe, DPI2 and above, have relatively high computational cost compared with efficiency of filtering power. With the second B-cell dataset, profiled using the Affymetrix U133 platform, there were 14 090 probe IDs, of which 1290 were TFs representing 1209 unique genes.

ARACNe identified 198 766 transcriptional interactions at the probe ID level. DPI2 produced 142 037 interactions, with
56 729 removed as indirect interactions. DPI3 produced 138 697 interactions, with 3340 additional ones removed as indirect interactions. Finally, DPI4 removed an additional 1533 interactions as indirect. We did not apply DPI of order greater than four.

In our comparative analyses, every TF had a different number of ARACNe- and hARACNe-inferred transcriptional targets. The distributions of transcriptional targets for ARACNe and hARACNe in the human B-cell analysis are shown in figure 2. For instance, from the U95 dataset, the ELK1 TF has 233 ARACNe-inferred targets, and we randomly selected ELK1 among the TFs which have more than 200 targets in ARACNe. Of these, 97 were removed by DPI2, leaving only 136 targets (figure 3). However, DPI3 removed only seven additional targets and no additional targets were removed by DPI4. An identical pattern was observed in the U133 human B-cell dataset analysis. Among all TFs, ZNF267 had the largest number of targets removed by higher-order DPI analysis. Of 223 ARACNe-inferred targets, 121 were removed by DPI2, and 30 and 5 additional targets were removed by DPI3 and DPI4, respectively, leaving only 86 targets.

As described, hARACNe removed a significant percentage of interactions inferred by ARACNe. This may improve the accuracy of inferred networks if the filtered interactions were enriched in false positives (i.e. indirect interactions). To test this hypothesis, we compared the ARACNe and hARACNe predictions against two datasets providing orthogonal evidence of TF–target interactions (see figure 2 for more details).

The first dataset was generated through ChIP-chip experiments designed to detect direct biochemical interactions between the MYC proto-oncogene and genome-wide promoter regions. The second dataset used RNAi experiments to detect genes differentially regulated upon inhibition of BCL6. GEPs were measured using the Affymetrix HG-U95A GeneChip platform with DLBCL and CLL cell lines.

### 3.2. ChIP-chip functional validation of MYC targets

We first compared ARACNe- and hARACNe-inferred targets of the MYC TF against genes with experimentally assessed interactions between the MYC protein and their proximal promoter regions. The assessment was based on existing ChIP-chip assays in Ramos and Mutu (human Burkitt’s lymphoma) cell lines [29]. Based on the CSA algorithm [28], 5307 and 3310 putative MYC-binding targets could be predicted in Ramos and Mutu cell lines, respectively. Comparing ARACNe- and hARACNe-inferred MYC regulons, from the U95Av2 data, 18 interactions were removed by hARACNe. Based on Ramos cell assays, 13 of these 18 interactions (hypergeometric test: \( p\)-value = 0.0178) were identified as false positives (i.e. no detected interaction between MYC and the gene’s promoter region) and only five were identified as true positives (table 3). Based on Mutu cell assays, 14 of the 18 interactions were identified as false positives, and only four of 18 were identified as true positives (hypergeometric test: \( p\)-value = 2.8713 x 10^{-7}).

As a result, it appears that hARACNe was three times more likely to remove false interactions not supported by a corresponding MYC-binding site in the proximal promoter region than to remove interactions with target whose promoter was bound by MYC.

### 3.3. Differentially expressed genes following shRNA-mediated BCL6 silencing

To further check whether the additional pruning produced by higher-order DPI improved the accuracy of TF–target prediction, we performed lentivirus-mediated shRNA silencing of the BCL6 gene and tested the enrichment of differentially expressed genes in ARACNe- versus hARACNe-inferred BCL6 targets. GEPs were measured using the Affymetrix HG-U95A GeneChip platform with DLBCL and CLL cell lines.
Where cells were infected with control non-target shRNA or validated shRNA targeting BCL6. Differentially expressed genes were identified using fold-change criteria and a 0.05 FDR threshold using a non-parametric U-test. In LY7, Pfeiffer and VAL cell lines, 1507, 3706 and 3199 differentially expressed genes were identified, respectively. From LY7 data, out of 334 ARACNe-inferred BCL6 targets, 318 were identified by sequential application of DPI2, DPI3 and DPI4. Twenty gains were achieved by hARACNe. Eighteen out of 20 were not differentially expressed following BCL6 silencing and may thus be considered false positives. Conversely, two BCL6 targets were newly identified due to consensus scoring analysis followed by 100 bootstrappings (hypergeometric test: p-value = 5.8955 × 10⁻¹⁴), which were differentially expressed following BCL6 knockdown and may have been considered true positives (table 4). In Pfeiffer cells, all 16 interactions removed by hARACNe(nested iterative procedures of DPI², DPI³ and DPI⁴) were identified as false positives and none were identified as true positives (hypergeometric test: p-value = 2.5294 × 10⁻⁸). Tables 3 and 4 also include the statistical analysis (Fisher’s exact test) with TPs and FPs from experimentally validated targets with each step of hARACNe-inferred targets in order to show how significantly targets were identified by either ARACNe or each cumulative order of hARACNe.

4. Discussion

The goal of this work was to improve the widely used ARACNe algorithm by further reducing false positive interactions, thus leading to more accurate inference of interaction networks. We note that our work is not intended to address other foundational issues of reverse engineering approaches or of the ARACNe algorithm, such as the assumptions inherent to the use of mRNA data. Rather, we use the widely accepted ARACNe assumptions as a starting point and assess the ability of higher-order DPI analysis to improve prediction accuracy based on theoretical arguments and comparison with independent ‘gold standard’ datasets. Within this limited context, we

![Figure 3](http://rsfs.royalsocietypublishing.org/)

**Figure 3.** Correlation distribution of the most differential TF targets between original ARACNe and hARACNe networks.
Table 3. (a) ChIP-chip experimental validation with Ramos and Mutu cell lines for MYC targets and (b) hypergeometric test in order to check if the gain is significantly derived from FPs.

<table>
<thead>
<tr>
<th>MYC: ChIP-chip</th>
<th>Ramos</th>
<th>Mutu</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>HG-U95Av2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DPI1</td>
<td>214</td>
<td>258</td>
</tr>
<tr>
<td>DPI2</td>
<td>210</td>
<td>249</td>
</tr>
<tr>
<td>DPI3</td>
<td>209</td>
<td>246</td>
</tr>
<tr>
<td>DPI4</td>
<td>209</td>
<td>245</td>
</tr>
<tr>
<td>gain</td>
<td>5</td>
<td>13</td>
</tr>
</tbody>
</table>

| HG-U95Av2      |       |      |                       |                                   |       |      |                       |                                   |
| total gain between DPI1 and DPI4 (nested) | Ramos | Mutu |
|               | TP    | FP   | hypergeometric test p-value |       | TP    | FP   | hypergeometric test p-value |
|               | 18    |      | 0.01779333             |       | 18    |      | 2.87 \times 10^{-4}      |
Table 4. (a) Gain and loss BCL6 targets comparison between ARACNe and hARACNe by shRNA library and (b) hypergeometric test in order to check if the gain is significantly derived from FPs.

### (a) Gain and loss BCL6 targets comparison between ARACNe and hARACNe by shRNA library

<table>
<thead>
<tr>
<th>BCL6: shRNA library</th>
<th>LY7</th>
<th>Fisher's exact test p-value of inferred interactions</th>
<th>Pfeiffer</th>
<th>Fisher's exact test p-value of inferred interactions</th>
<th>VAL</th>
<th>Fisher's exact test p-value of inferred interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>HG-U95Av2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DPI1</td>
<td>54</td>
<td>280</td>
<td>0.0136</td>
<td>144</td>
<td>190</td>
<td>1.06 × 10⁻⁷</td>
</tr>
<tr>
<td>DPI2</td>
<td>57</td>
<td>265</td>
<td>0.001614</td>
<td>146</td>
<td>176</td>
<td>1.60 × 10⁻⁹</td>
</tr>
<tr>
<td>DPI3</td>
<td>56</td>
<td>262</td>
<td>0.001988</td>
<td>144</td>
<td>174</td>
<td>2.28 × 10⁻⁹</td>
</tr>
<tr>
<td>DPI4</td>
<td>56</td>
<td>262</td>
<td>0.001988</td>
<td>144</td>
<td>174</td>
<td>2.28 × 10⁻⁹</td>
</tr>
<tr>
<td>gain</td>
<td>2</td>
<td>18</td>
<td>n.a.</td>
<td>0</td>
<td>16</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

### (b) Hypergeometric test for total gain between DPI1 and DPI4 (nested)

<table>
<thead>
<tr>
<th>HG-U95Av2</th>
<th>LY7</th>
<th>hypergeometric test p-value</th>
<th>Pfeiffer</th>
<th>hypergeometric test p-value</th>
<th>VAL</th>
<th>hypergeometric test p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
<td>2</td>
<td>18</td>
<td>16</td>
<td></td>
<td>16</td>
</tr>
</tbody>
</table>

gain 2 18 n.a. gain 0 16 n.a. gain 0 16 n.a.
believe that this work represents a useful contribution to the field, based on the widespread use of ARACNe and the benefit of an extension improving its accuracy.

From a theoretical standpoint, we note that higher-order, indirect Markov chain interactions detected by hARACNe would also be eliminated by repeated application of the DPI, under certain assumptions. Specifically, these would require that (i) the network have a tree (or locally tree-like) structure, (ii) the network contain only pairwise interactions, and (iii) mutual information be measured without errors. Given that each of these assumptions may be individually violated and given the large number of potential pairwise interactions, higher-order DPIs provide additional filters to increase the global Markov chain stringency in ARACNe-inferred networks, thus eliminating indirect interactions that may have been missed by DPI1. In this context, the experimental data we have provided, demonstrating a higher accuracy of hARACNe networks, represents the most meaningful test of the method.

Yet, there are a few computational limitations in hARACNe. Firstly, hARACNe starts from an ARACNe-inferred network, on which it nested iteratively applies higher-order DPI analyses, it must consider a large number of candidate Markov chain paths, traversing up to four interactions, and is thus computationally intensive. Our results, however, show that DPI2 provides the greatest pruning effect while higher-order DPIs have a significantly lower detection rate. Thus, one may want to consider whether to apply only DPI1 or up to DPI3 to obtain the highest increase in accuracy at the lowest computational cost.

Our experimental design for evaluating hARACNe relied on two datasets providing orthogonal evidence of direct TF–target interactions. Specifically, we used ChIP-chip assays and shRNA-mediated silencing to show that hARACNe preferentially removes false positive interactions. In BCL6 silencing experiments, interactions removed by hARACNe had significantly more false positives than true positives. Similarly, when considering ChIP-chip experimental data on MYC-binding sites, a similar improvement in the false positive predictions was introduced by hARACNe. Overall, the higher-order DPI analysis inferred a more reliable network, compared with the DPI1 procedure of the original ARACNe algorithm.

This project was partly supported by grant no. U54CA149237 from the Integrative Cancer Biology Program of the National Cancer Institute and National Institutes of Health grant no. U54CA121852 (National Centers for Biomedical Computing).

References


