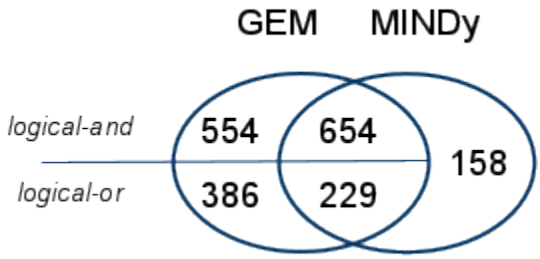
Comparison of GEM results with MINDy

Özgün Babur, Emek Demir, Mithat Gönen, Chris Sander, and Ugur Dogrusoz
Supplementary Document for the article "Discovering Modulators of Gene Expression"

MINDy (1), similar to GEM, uses gene expression profiles to predict modulators of transcription factors. MINDy uses a different metric, namely, differential conditional mutual information ($\Delta\mathrm{CMI}$) to estimate the effect of interaction between factor and modulator on the target's expression level.

In order to compare the results of GEM and MINDy, we applied both methods to detect the modulators of AR using the expression dataset of Expression Project for Oncology (expO). We tested 70 modulator candidates (interactors of AR with variance greater than 1) using 47 target genes. We combinatorially created 34270 triplets to test (one gene can have multiple rows in the expression data; each row is tested separately).

Using a 1% false discovery rate, GEM detects 1823 triplets with significant γ. MINDy detects 1041 triplets with significant $\Delta\mathrm{CMI}$. 883 of these triplets are the same. 34% of detected GEM triplets are *logical-or* cases, and filtered out leaving 1208 triplets in the results. MINDy does not differentiate between *logical-or* and *logical-and*, and reports all result triplets as modulation. However, we see that 26% of the common result triplets are detected as *logical-or* by GEM.



MINDy can infer 54% of GEM results, while GEM can infer 85% of MINDy results.

GEM classifies modulations into 6 classes as described in the section "Category of action". MINDy classifies biological activity of modulators into 2 classes - *activator* and *antagonist* - according to the following function:

$$class = \begin{cases} activator & if \rho(\mu_t^+ - \mu_t^-) > 0 \\ antagonist & if \rho(\mu_t^+ - \mu_t^-) < 0 \\ undetermined & otherwise \end{cases}$$

Where ρ is the Pearson correlation between F and T, μ_t^+ is the mean expression of T when M is high, and μ_t^- is the mean expression of T when M is low. MINDy leaves the class undetermined if the difference between means is not significant. Conceptually, "Enhances Activation" and "Enhances Inhibition" classes of GEM correspond to *activator* class of MINDy; and "Attenuates Activation" and "Attenuates Inhibition" classes of GEM correspond to *antagonist* class of MINDy.

Below table shows a detailed intersection of GEM and MINDy results. Row headers are GEM

classes, and column headers are MINDy classes.

GEM / MINDy	Activator	Antagonist	Undetermined	Undetected	Total	MINDy det.
Enhances Activation	139	40	33	67	279	76%
Enhances Inhibition	78	19	49	96	242	60%
Attenuates Activation	22	72	28	162	284	43%
Attenuates Inhibition	22	46	14	105	187	44%
Inverts Activation	25	25	2	59	111	47%
Inverts Inhibition	14	0	4	38	56	32%
Unclassified	0	12	10	27	49	45%
Logical-or	69	147	13	386	615	
Undetected	68	48	42	32289	32447	
Total	437	409	195	33229	34270	
GEM detection %	84%	88%	78%			

As we expected, most of the "Enhances ..." classes of GEM are categorized as *activator* by MINDy, and most of the "Attenuates ..." classes of GEM are categorized as *antagonist* by MINDy. However, there are substantial number of triplets in conflict (24%); i.e., enhancers are categorized as antagonists, and attenuators are categorized as activators.

To better understand the source of these conflicts, consider the triplet M = DDC, F = AR, and T = LIFR, whose observed proportions are given below:

M \ F	-	+
-	0.50	0.54
+	0.11	0.80

$$\alpha_f = 0.04, \quad \alpha_m = -0.39, \quad \beta_f = 0.69, \quad \beta_m = 0.26$$

Proportions here indicate when M is high, F upregulates T more often. Effect of F in M- (α_f) is insignificant, but F strongly activates T in M+ ($\beta_f > 0$), M activates T in F+ ($\beta_m > 0$), and combined effect of M and F is again an activation ($\alpha_f + \beta_m > 0$). This is a typical GEM class *Enhances Activation*. We also observe that M has a negative effect on T in F- ($\alpha_m < 0$), which is probably not related to the modulation of F.

On the MINDy side, correlation between F and T (ρ) is 0.29, and the difference of mean T expression in M+ and M- ($\mu_t^+ - \mu_t^-$) is -0.53. So, MINDy classifies this triplet as *antagonist*.

Here, we see that MINDy is assuming a positive $\mu_t^+ - \mu_t^-$, when the modulator is an

activator and F-T correlation is positive. However, this assumption does not hold in this case since the effect of M in low F causes a negative shift in T expression.

Causes of other conflicts are similar. They are related to a questionable assumption by MINDy that disregards the low F-dependent effect of M.

References

1. Wang,K., Saito,M., Bisikirska,B.C., Alvarez,M.J., Lim,W.K., Rajbhandari,P., Shen,Q., Nemenman,I., Basso,K., Margolin,A.A., Klein,U., Dalla-Favera,R. and Califano,A. (2009) Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat Biotechnol.,* **27**, 829–839.