

# Network Expansion and Pathway Enrichment Analysis towards Biologically Significant Findings from Microarrays

Xiaogang Wu<sup>1,2\*</sup>, Hui Huang<sup>1,2\*</sup>, Tao Wei<sup>3</sup>, Ragini Pandey<sup>2</sup>, Christopher Reinhard<sup>3</sup>, Shuyu D. Li<sup>3§</sup>, Jake Y. Chen<sup>1,2§</sup>

<sup>1</sup> School of Informatics, Indiana University, Indianapolis, IN 46202

<sup>2</sup> Indiana Center for Systems Biology and Personalized Medicine, Indiana University, Indianapolis, IN 46202

<sup>3</sup> Eli Lilly and Company, Indianapolis, IN 46285, USA

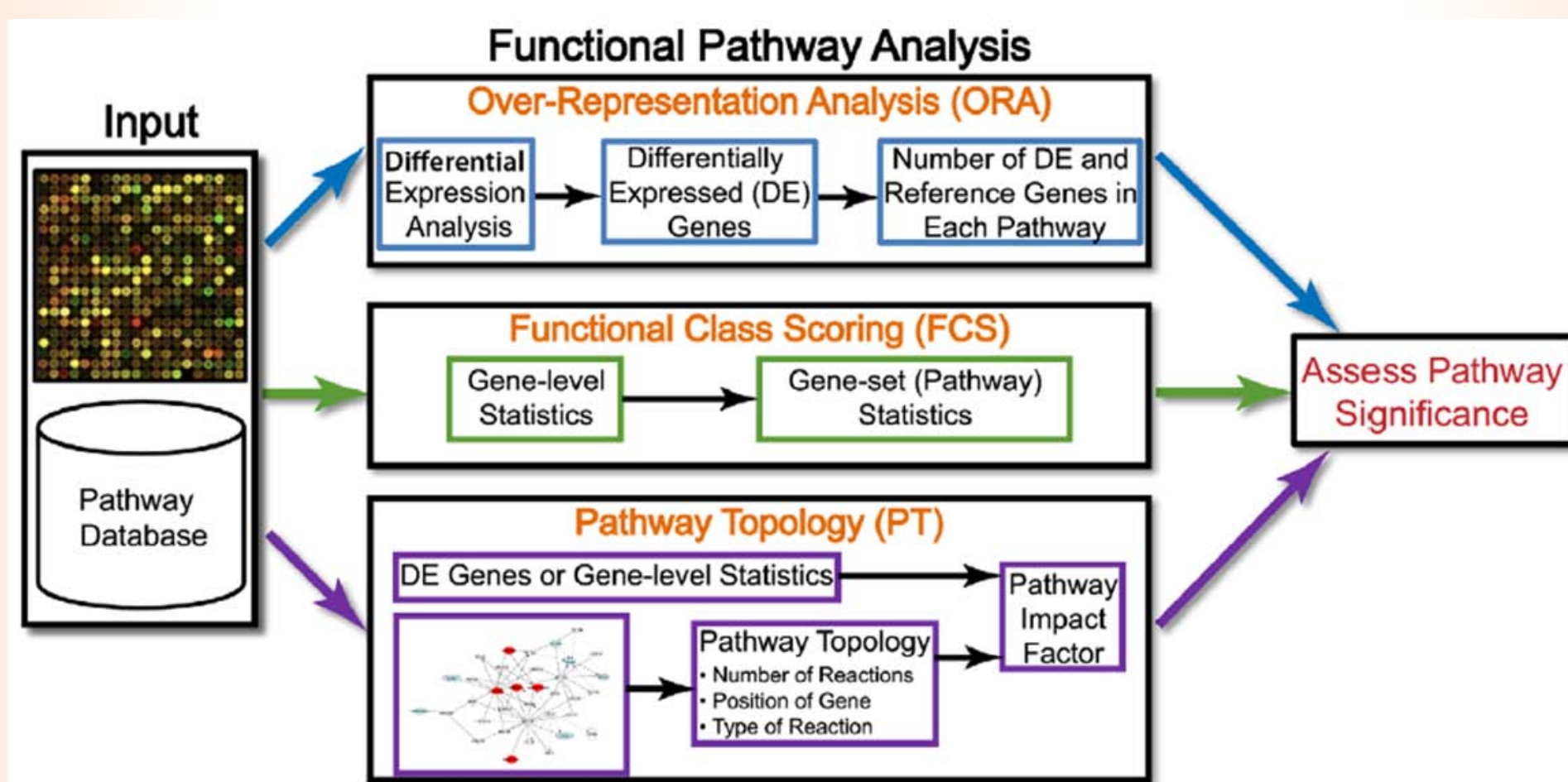
\* Email: [jakechen@iupui.edu](mailto:jakechen@iupui.edu) Website: <http://bio.informatics.iupui.edu/> Phone: (317) 278-7604



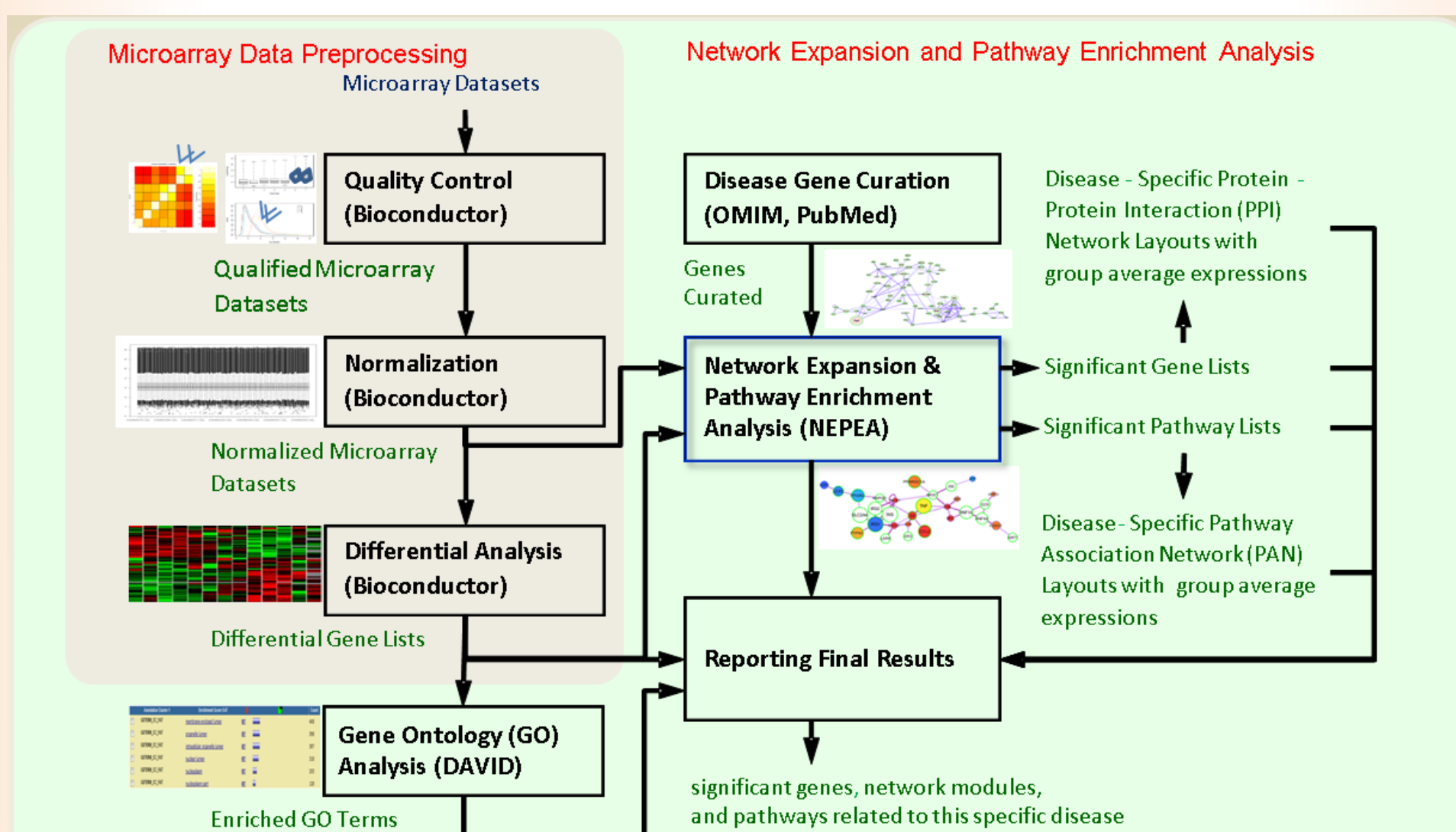
## Introduction

In many cases, crucial genes show relatively slight changes between groups of samples (e.g. normal vs. disease), and many genes selected from microarray differential analysis by measuring the expression level statistically are also poorly annotated and lack of biological significance.

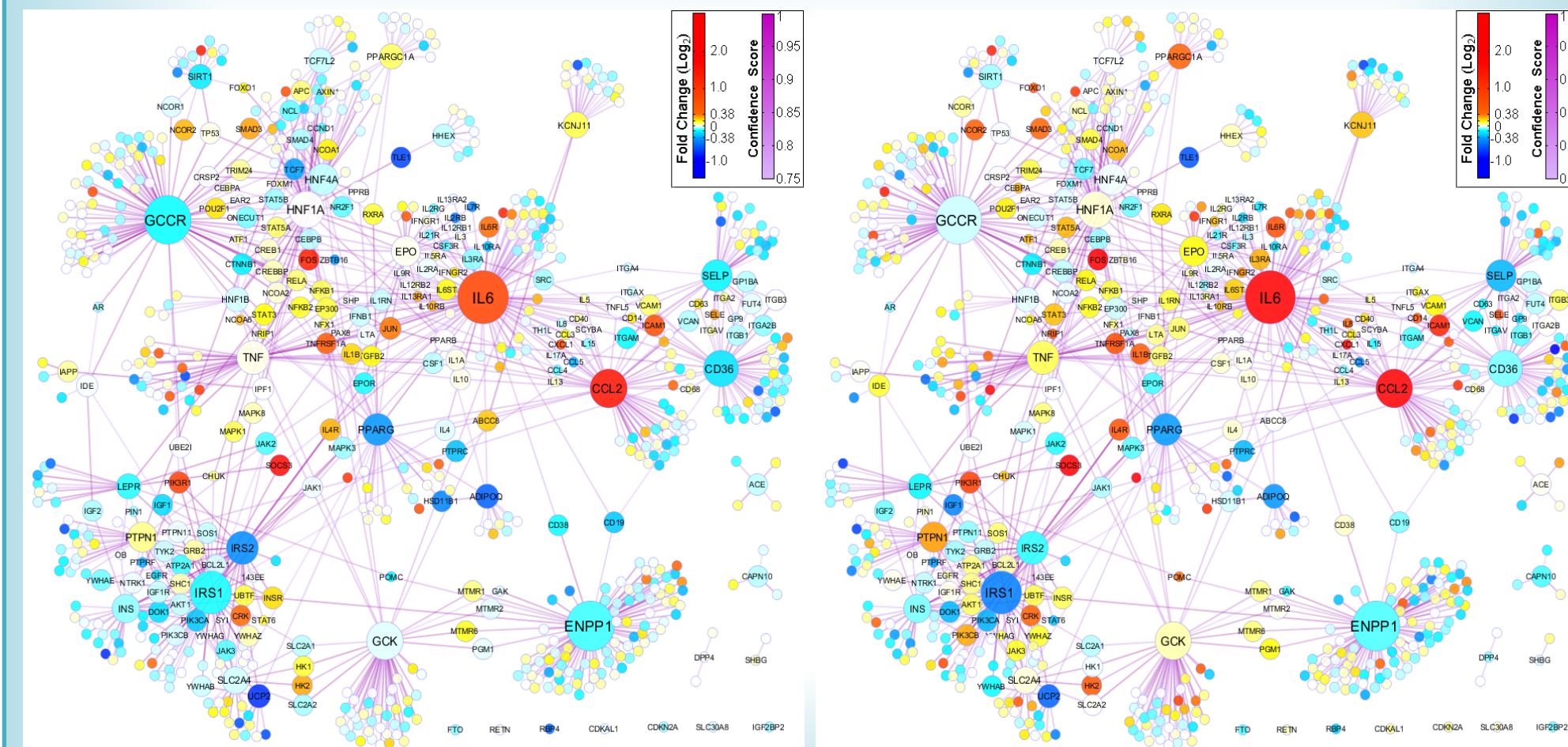
In this work, we present an innovative approach - network expansion and pathway enrichment analysis (**NEPEA**) for integrative microarray analysis. We assume that organized knowledge will help microarray data analysis in significant ways, and the organized knowledge could be represented as molecular interaction networks or biological pathways.



## Framework

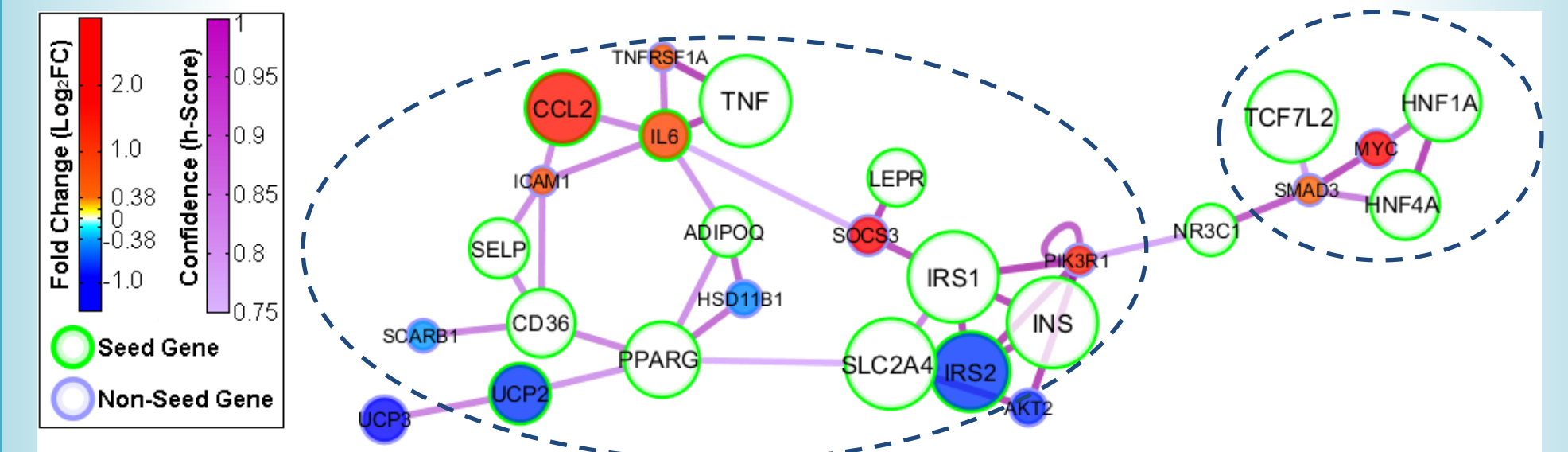


## T2D-specific PPI Networks (GSE24215: IBB vs. IAB)

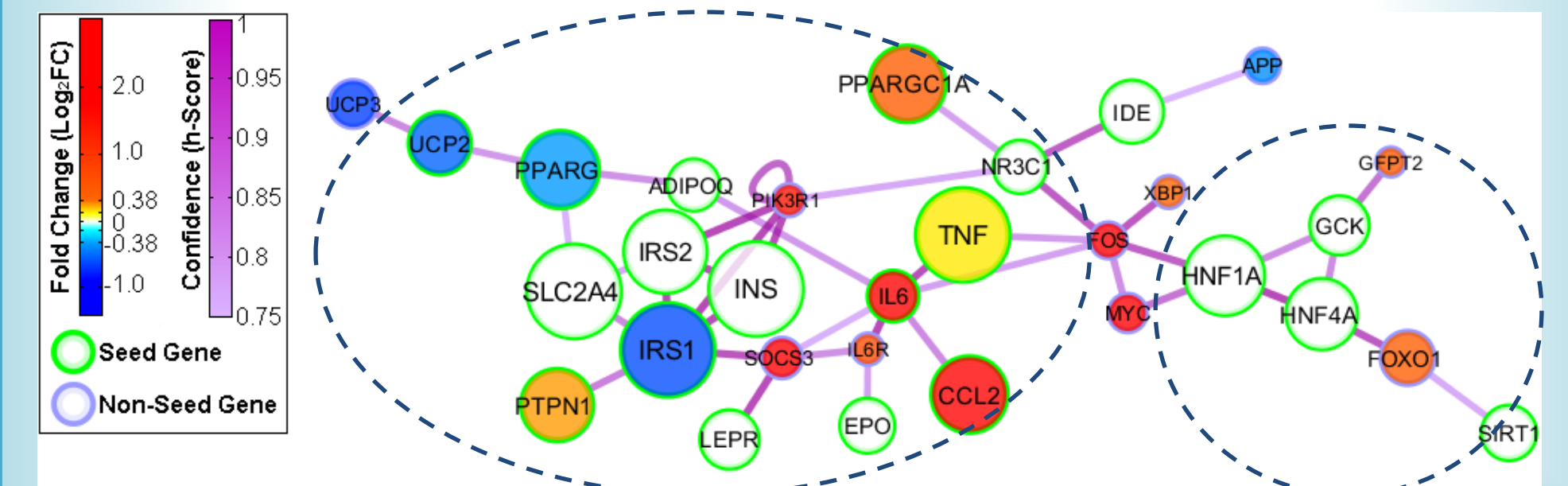


## Top Biologically-Significant Gene Network

### GSE24215 Muscle Tissue Dataset: IBB Group



### GSE24215 Muscle Tissue Dataset: IAB Group



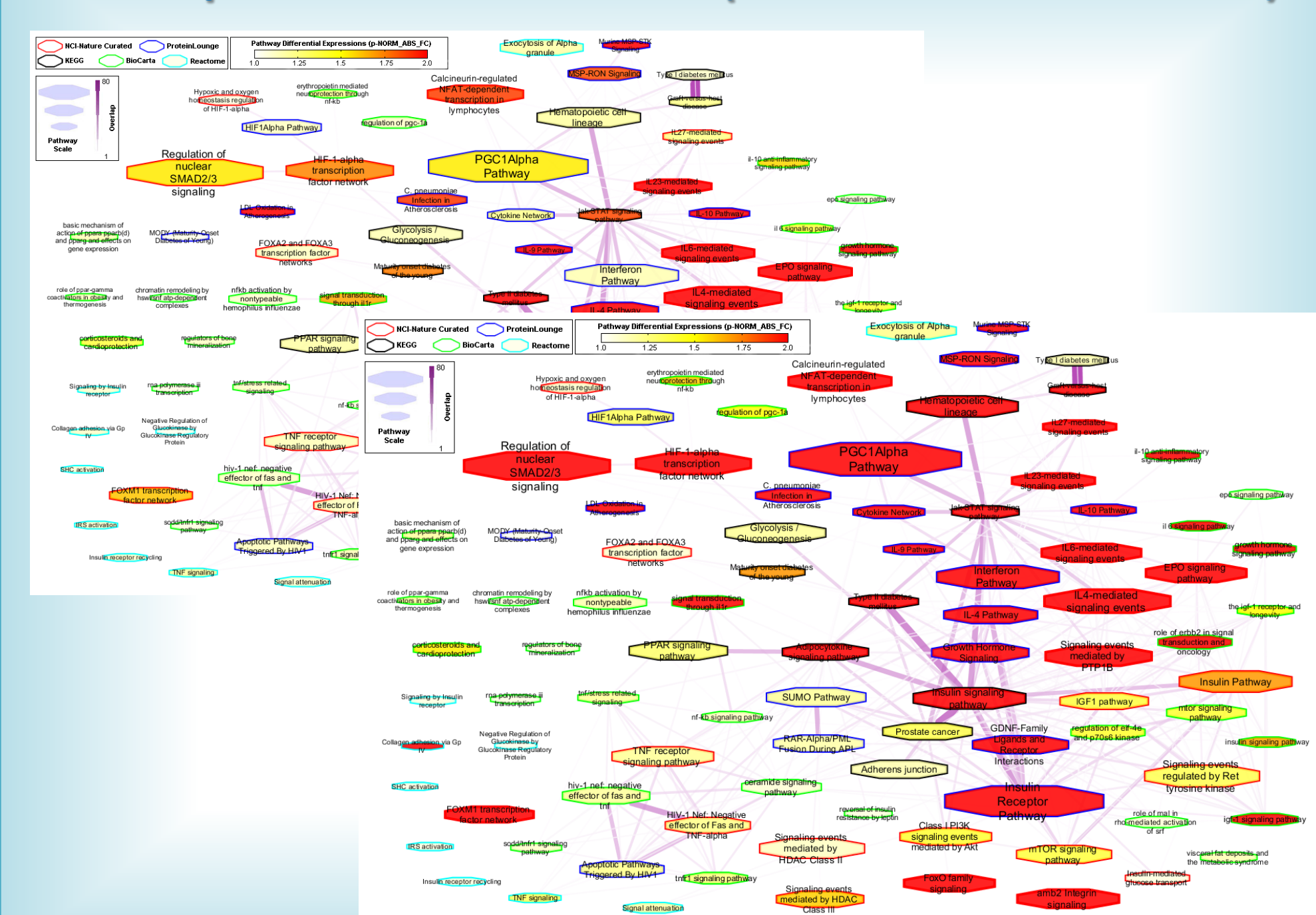
## Conclusions

- From the evidence from literature (PubMed), Top 20 significant genes from our analysis have more supports than Top 20 differential genes from simple differential analysis, in the case study on the microarray dataset - GSE24215.
- Our findings are not only consistent with the original findings mostly, but also obtained more supports from other literatures. This implies the vitality of our hypothesis on which organized knowledge will help microarray data analysis in significant ways.

## Top Biologically-Significant Genes (GSE24215\_IAB)

Gene Symbol	p-Value	FDR	Log2_FC	ABS_FC	Weight_1	Weight_2	Sig_Score	Evidences
CCL2	0.00013	0.01462	2.59339	6.03513	98	75.2645	34.9751	111
IL6	0.00044	0.03051	2.0786	4.22398	140	112.808	34.68919	52
IRS1	0.00902	0.16494	-0.68912	1.6123	103	83.1045	23.58864	280
IL6R	0.00026	0.00404	0.79678	1.73722	70	55.9728	22.14359	7
VEGFA	0.00017	0.01724	0.92831	1.90304	57	44.6818	21.6589	28
APP	0.03329	0.31458	-0.43503	1.35194	80	68.3971	21.01141	15
SOCS3	0.00005	0.0078	4.12982	17.50651	7	6.3574	20.52815	28
ADRB2	0.00016	0.0164	1.0105	2.01461	37	31.2254	20.09311	10
FOXO1	0.00008	0.01056	0.52622	1.44015	65	44.529	19.42493	59
MYC	0.00026	0.02179	2.86543	7.28754	10	8.8236	19.33394	23
SOD2	0.00164	0.06478	0.82246	1.76841	50	29.5486	18.85629	4
DGKD	0.02661	0.28238	0.5992	1.51487	42	34.932	18.59775	2
FOS	0.00031	0.02436	3.31416	9.9463	7	6.3552	18.17698	18
PIK3R1	<1.0E-5	0.00008	1.67218	3.18695	13	11.8228	17.1966	9
XBP1	0.00338	0.09606	0.40352	1.32273	42	26.9454	16.35234	9
AGT	0.00295	0.08852	0.59516	1.51064	24	18.032	15.28071	42
UCP3	0.00163	0.06449	-0.75312	1.68543	22	15.2626	15.10061	45
UCP2	0.00186	0.06897	-0.60181	1.51762	22	15.7096	14.63273	78
PPARGC1A	0.00479	0.117	0.53018	1.44411	17	13.6	13.65437	111
GFPT2	0.04473	0.36274	0.54879	1.46286	14	12.2852	13.2432	2

## T2D-specific PAN Networks (GSE24215 : IBB vs. IAB)



Supported by

