

Employing Publicly Available Biological Expert Knowledge from Protein- Protein Interaction Information

Kristine A. Pattin, Ph.D.

Dartmouth Medical School

Hanover, NH USA

Sept. 24, 2010

Kristine A. Pattin, Jiang Gui, and Jason H. Moore

Pattern Recognition in Bioinformatics

Nijmegen, The Netherlands

Computing Time Required

(assuming 1 million PCs each processing 1 model/sec.)

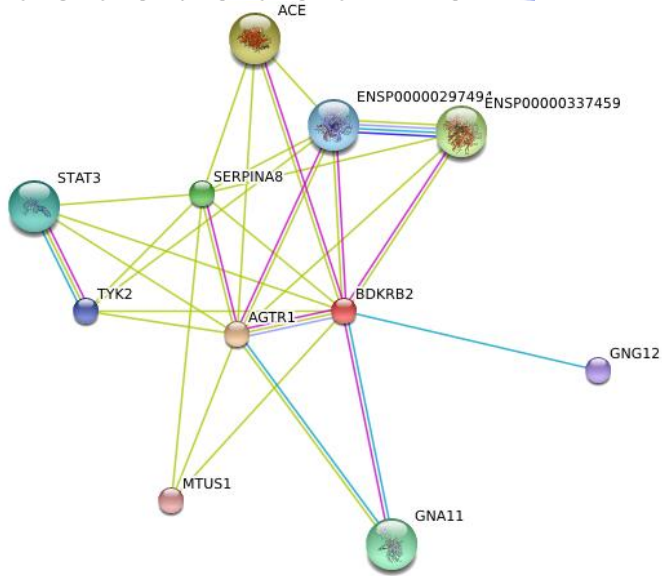
Test all subsets of 1 SNP:	$n = 3 * 10^5$
Test all subsets of 2 SNPs:	$n \approx 4 * 10^{10}$
Test all subsets of 3 SNPs:	$n \approx 4 * 10^{15}$
Test all subsets of 4 SNPs:	$n \approx 3 * 10^{20}$

Test all subsets of 1 SNP:	1 second
Test all subsets of 2 SNPs:	11 hours
Test all subsets of 3 SNPs:	127 years
Test all subsets of 4 SNPs:	9513 millennia

Neighborhood
 Gene Fusion
 Cooccurrence
 Coexpression
 Experiments
 Databases
 Textmining
 [Homology]

Score

			●	●	●	0.999
			●	●	●	0.999
			●	●	●	0.999
			●	●	●	0.999
			●	●	●	0.999
			●	●	●	0.998



STRING

- Over 2.5 million interactions
- Incorporates a number of databases and information sources
- Confidence Score
- Metrics: AVE, SUM, MAX-AVE, MAX-SUM

This is the **evidence view**. Different line colors represent the types of evidence for the association



Results

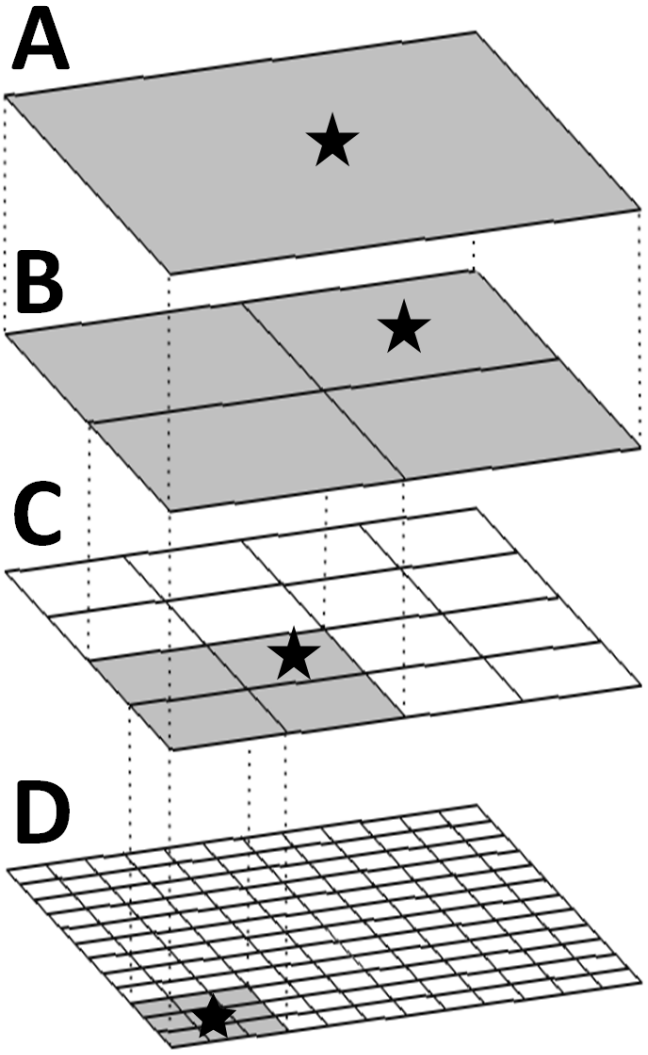
	AVE	SUM	MAX-SUM	MAX-AVE	Spearman P-value
AVE					0.40
SUM	0.016*				0.007*
MAX-SUM	0.031*	0.160			0.05*
MAX-AVE	0.031*	0.290	0.030*		0.05*

*AVE > SUM, MAXSUM, MAXAVE

*MAXSUM < AVE, MAXAVE

*MAXAVE < AVE

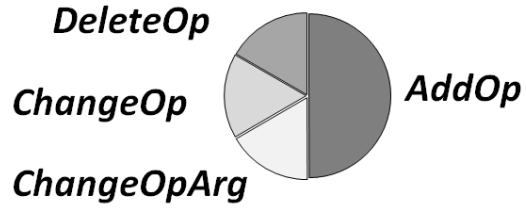
DNA Repair Gene Interactions	AVE	SUM	MAX-AVE	MAX-SUM
XRCC1.XRCC3	288	346	1126	1176
XRCC1.XPD	288	234	1130	1158



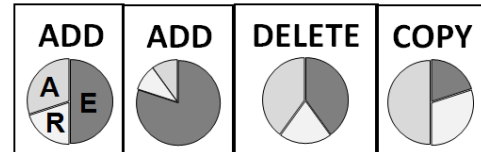
Mutation Probability

$P = 0.1$

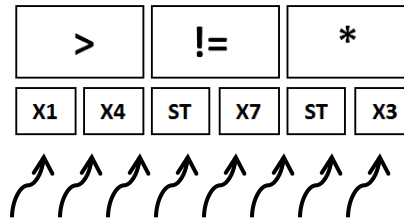
Mutation Operator



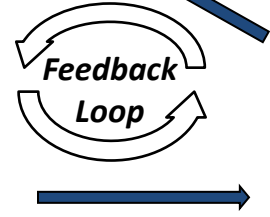
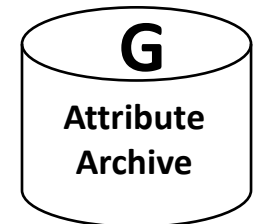
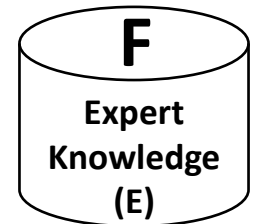
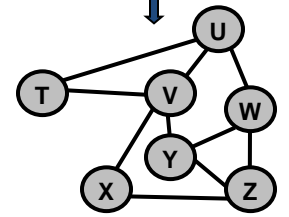
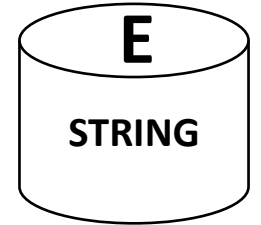
Solution Operator



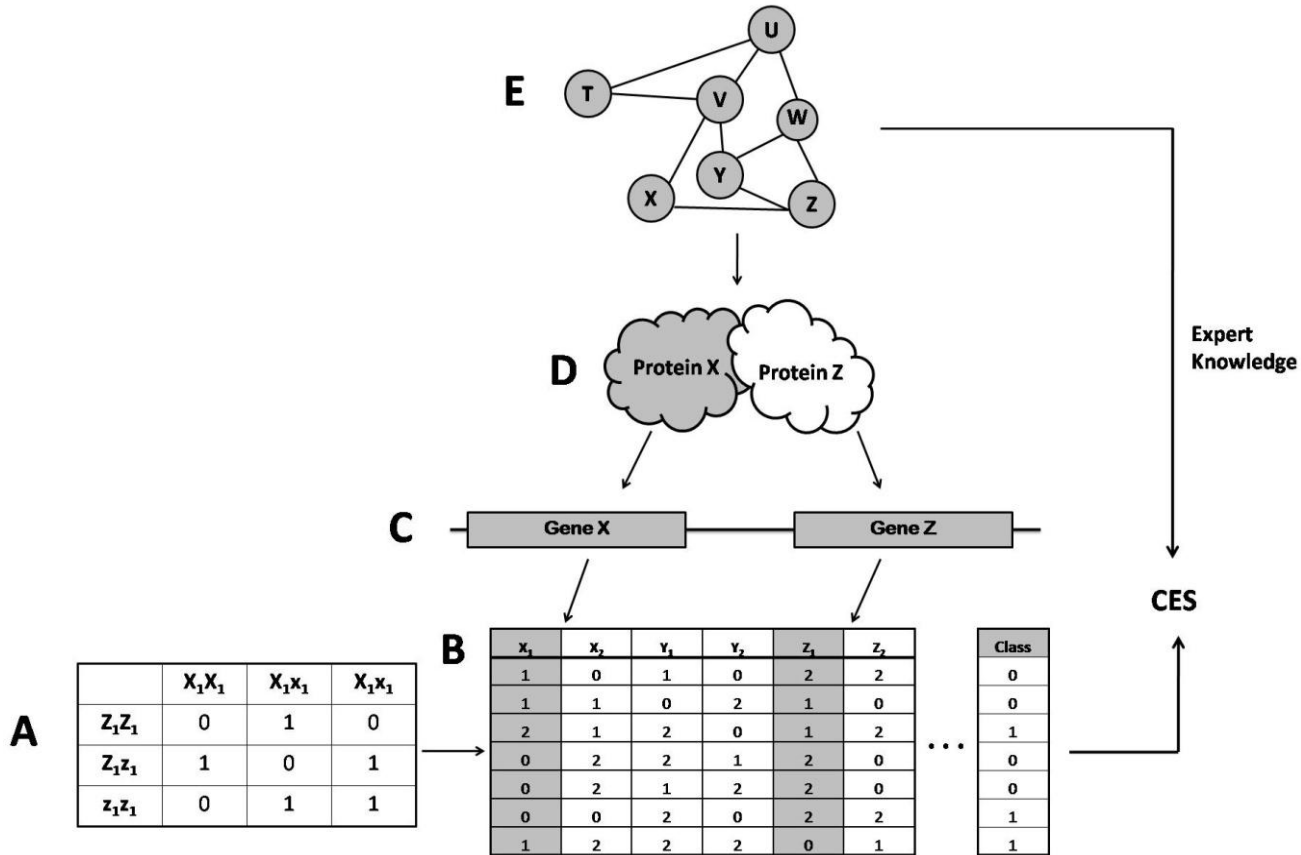
Solution



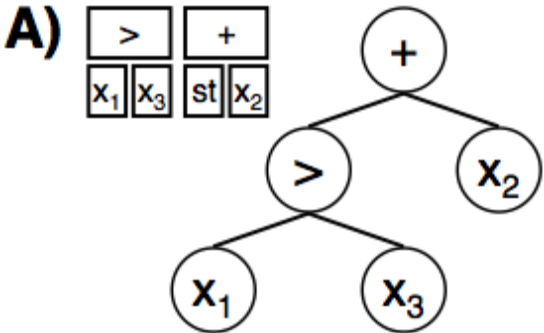
H Environmental Noise



Incorporating Biological Expert Knowledge in CES

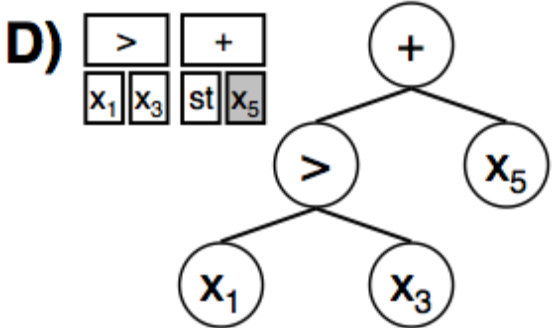
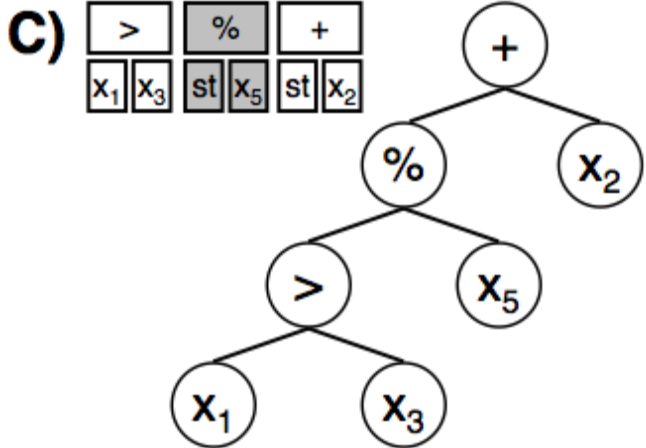


Solution Operators and Biological Expert Knowledge



B)

	x ₁	x ₂	x ₃	x ₄	x ₅
x ₁	0	.71	.84	.10	0
x ₂	.63	0	.68	0	0
x ₃	.84	.71	0	.10	.97
x ₄	.24	0	.10	0	.32
x ₅	0	0	.81	.32	0



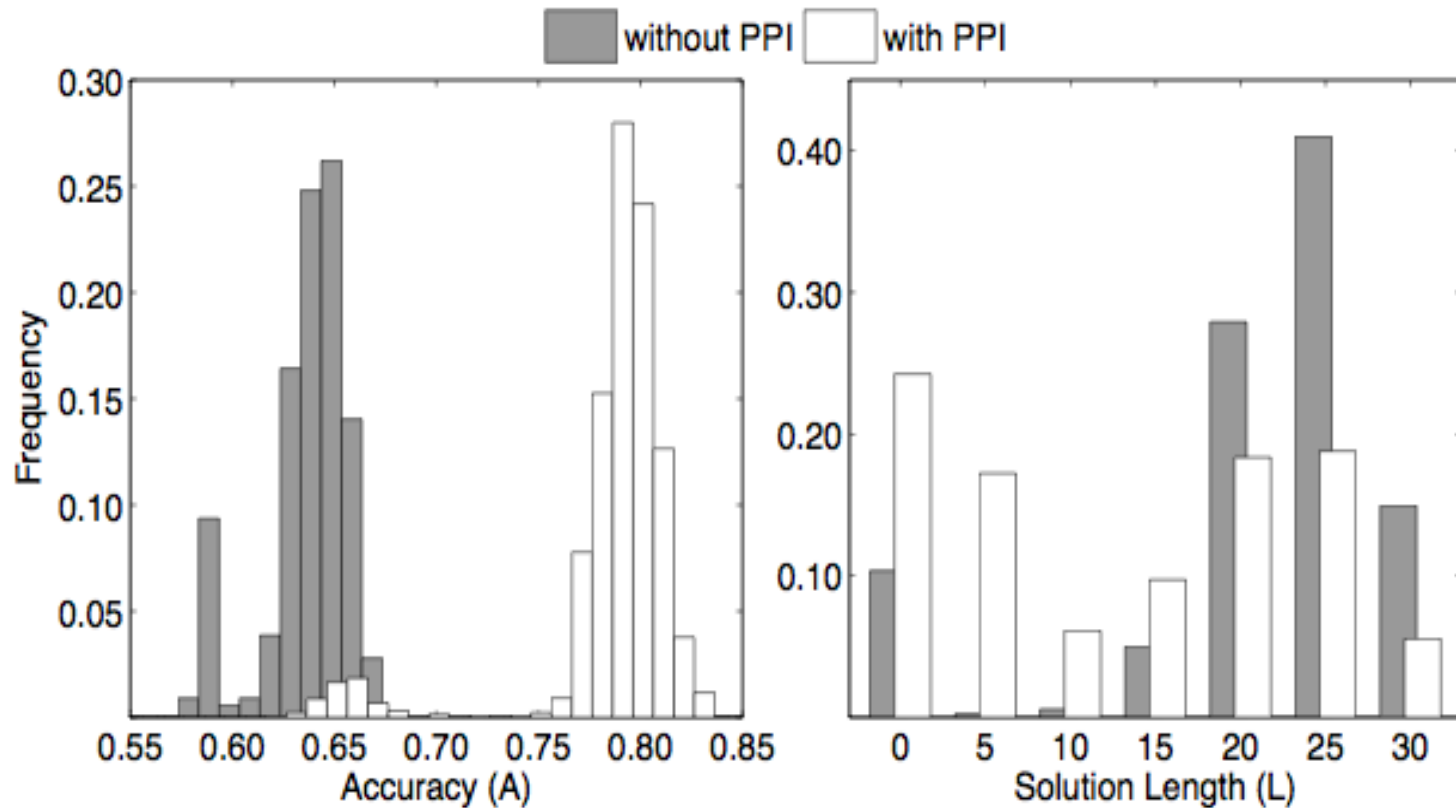
ADD

ALTER

Results

PPI	Confidence score of PPI	With expert knowledge	Control
CASP3-CASP9	0.998	100%	0%
CASP3- TNFRSF1A	0.963	100%	0%
CASP3-AKT1	0.933	96%	0%
CASP3-TNF	0.916	80%	0%

Solution Accuracy and Length



Conclusion

- CES exhibits the ability to indentify simulated epistatic interactions in more concise models of disease susceptibility
- Accuracy is improved using PPI biological expert knowledge
- CES is sensitive to PPI confidence

- CES may be ready to tackle a wide range of real world data
- Real data brings its own unique challenges
- Analysis of real world data will require giving CES access to other sources of knowledge
 - Biochemical pathway information
 - Gene Ontology (GO)
 - Chromosomal location
 - Statistical
- Future work will focus on the application of this approach to the genetic analysis of human disease