



MOUNT SINAI  
SCHOOL OF  
MEDICINE



# Construction and Analysis of a Mammalian Kinase-Kinase Regulatory Network

Mariola K. Szenk<sup>1</sup>, Alexander Lachmann<sup>2</sup>, Shan Zhao<sup>2</sup>, Avi Ma'ayan<sup>2</sup>

<sup>1</sup>Macaulay Honors College at Hunter College,

<sup>2</sup>Systems Biology Center New York and Department of Pharmacology and Systems Therapeutics, Mount Sinai School of Medicine, New York NY

MACAULAY  
HONORS COLLEGE AT CUNY

HUNTER  
The City University of New York

## ABSTRACT

Kinases are enzymes that transfer phosphate groups to their substrates. They phosphorylate linear motifs on Serine, Threonine, or Tyrosine amino-acid residues. Phosphorylation regulates protein-protein interactions, protein translocation, protein degradation and protein enzymatic activity of kinase substrates. These events play vital roles in all cellular regulatory processes. The sequencing of the human genome identified 518 kinase genes. Since many protein kinases are substrates for other kinases and such relationships are known, we constructed a kinase-kinase mammalian regulatory network from data reporting kinase-substrate relationships from the literature. This network contains 385 kinases linked by 2156 directed edges. Using the Markov Clustering and the Molecular Complex Detection Algorithms, we detected clusters of protein kinases in this network. Such clusters are subgraphs of densely interconnected nodules of kinases which may have specific functional regulatory roles. To associate function to clusters we performed statistical enrichment analysis of the kinase clusters against background datasets of lists of genes with a common associated function. We compared the kinase clusters against OMIM (Online Mendelian Inheritance in Man), GO (Gene Ontology), and protein domains from InterPro and PFAM to yield a better perspective of the kinase network by understanding the relationships between and within the kinase clusters.

## METHODS

### CLUSTERING

Clusters are highly interconnected regions in networks where nodes within clusters often have common functions. Clustering algorithms allows for the detection of modules within large networks. Clustering algorithms automatically detect clusters, which in our case are groups of protein kinases identified for further study.

### MOLECULAR COMPLEX DETECTION ALGORITHM

```

Stage 1: Vertex Weighting
procedure MCODE-VERTEX-WEIGHTING
input graph, G = (V,E)
for all v in G do
  N = find neighbors of v to depth 1
  A = Get highest score graph from N
  n = Get highest score number from N
  Set weight of v = A * n
end procedure

Stage 2: Molecular Complex Prediction
procedure MCODE-FIND-COMPLEX
input graph, G = (V,E), vertex weights, W;
vertex weight percentage, P; seed vertex, v;
if v already seen then return
for all neighbors of v do
  if weight of v >= (weight of v) * (1 - P) then add v to complex C
  call MCODE-FIND-COMPLEX (G, W, v)
end for
end procedure

procedure MCODE-FIND-COMPLEXES
input graph, G = (V,E), vertex weights, W;
vertex weight percentage, P;
for all v in G do
  Find neighbors of v then call MCODE-FIND-COMPLEX (G, W, v)
end for
end procedure
  
```

Fig 2. Pseudocode of the MCODE clustering algorithm

### MARKOV CLUSTERING ALGORITHM

The MCL algorithm simulates random walks in a network, reassigning probability values associated with the transition from one node to another within the graph. MCL works on the premise that clusters will be connected by paths that are more frequently visited. Each MCL iteration alternates between expansion and inflation of the graph's adjacency matrix before producing an output (Fig 3).

#### Expansion phase

The matrix is expanded by calculating the linear algebraic matrix-matrix multiplication of the original matrix.

#### Inflation phase

Each non-zero value in the matrix is raised to a power, such that all values below 1 become smaller and those larger than 1 become larger. After diagonal scaling of the resultant matrix, values below a certain threshold are dropped after the normalization (scaling) step in each iteration.

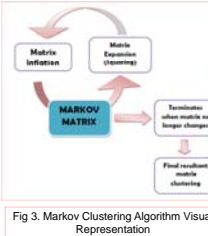


Fig 3. Markov Clustering Algorithm Visual Representation

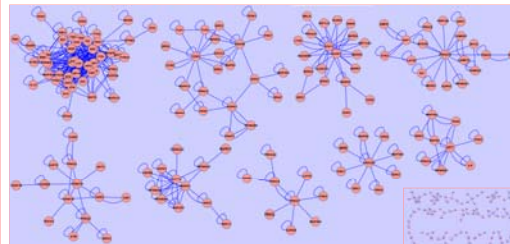


Fig 4 Kinase-kinase clusters identified using MCL

## KINASE REGULATORY NETWORK

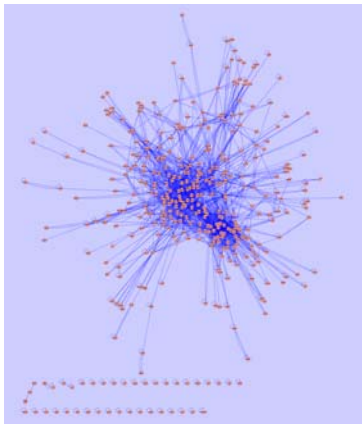


Fig 1. A Mammalian Kinase-Kinase Network (385 nodes, 2156 edges, 259 self-loops) was created from publicly available databases (NetworkKIM, Phospho.ELM, MINT, HPRD, PhosphoPoint, IMB, and Swiss-Prot ) reporting experimentally verified kinase-substrate interactions. The kinase-kinase network is visualized using Cytoscape.

## LISTS2NETWORKS

Lists2Networks is a web-based tool developed at the Ma'ayan lab to allows users to upload lists of mammalian genes and cross-reference them against background gene lists in other databases. L2N also provides a network view of lists and their similarities with other lists. Statistical gene-enrichment analyses were used to determine relationships within and between our clusters of kinases.

### FISHER'S EXACT TEST WITH THE BONFERRONI CORRECTION

L2N uses the Fisher's Exact Test to find gene-sets with unexpected significant overlap against background lists. This is a statistical comparison test.

Observed states	a	c	R <sub>1</sub>
Observed states	b	d	R <sub>2</sub>
	C <sub>1</sub>	C <sub>2</sub>	N

$$N = \sum_{i=1}^r \sum_{j=1}^c C_{ij}$$

$$P_{\text{null}} = \frac{(R_1! R_2! \dots R_m!) (C_1! C_2! \dots C_n!)}{N! \prod_i a_{ij}!}$$

Fig 5. The formula for determining p-values

The Bonferroni correction is applied to adjust p-values for large scale comparisons. To adjust using the Bonferroni correction, the p-value is multiplied by the number of outcomes being tested. If the adjusted p-value is greater than 1.0, it is rounded down to 1.0.

## RESULTS

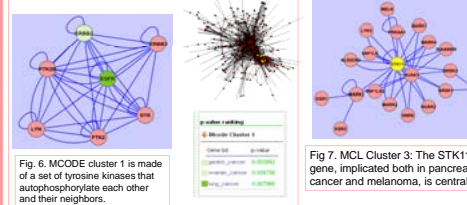


Fig 6. MCODE cluster 1 is made of a set of tyrosine kinases that autophosphorylate each other and their neighbors.

Fig 7. MCL Cluster 3: The STK11 gene, implicated both in pancreatic cancer and melanoma, is central.

Different kinase-kinase regulatory clusters play different roles in known cell signaling pathways. To associate the kinase clusters with pathways we used WikiPathways to compute statistically significant overlap between known pathways and identified kinase clusters

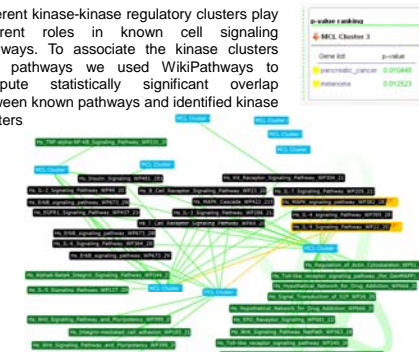


Fig 8. Signaling pathways from WikiPathways sharing kinases found in clusters identified using MCL

## TYROSINE KINASES IMPLICATED IN CANCER PATHOPHYSIOLOGY

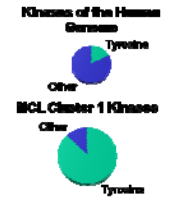
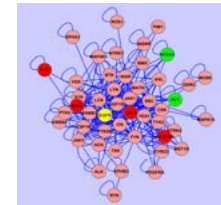


Fig 9. MCL Cluster 1 is a "cancer" cluster

**GO ENRICHMENT:**

- Mesoderm Development (middle germ layer that develops into muscle, bone, cartilage, blood and connective tissue)
- Positive Regulation of DNA Replication
- Hemopoiesis (maturation of the myeloid and lymphoid derived organ/tissue systems)

Phosphorylation sites serve as switches for exposing structural domains for binding with other proteins. Protein structural domain enrichment analysis shows that different kinase clusters utilize selective structural domains.

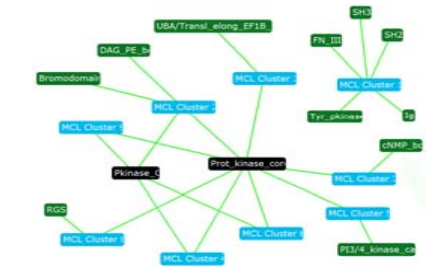


Fig 10. Network view of kinase clusters and their structural domain associations.

## BIBLIOGRAPHY

1. Bader, G., & Hogue, C. *BMC Bioinformatics*, 4: 1 (2003)
2. Enright, A., Dongen, S. V., & Ouzounis, C. *Nucleic Acids Res*, 30:1575-1584 (2002)
3. Lachmann, A., & Ma'ayan, A. *Bioinformatics*, 25:684-686 (2009)
4. Manning, G., Whyte, D., Martinez, R., Hunter, T., & Sudarsanam, S. *Science*, 298:1912 (2002)
5. Tan, C. S., Bodenmiller, B., Pasculescu, A., Jovanovic, M., Hengartner, M., Jorgensen, C., Bader, G., Aebersold, R., Pawson, T., & Linding, R. *Science Signaling*, 2:1 (2009)
6. Tilford, C. A., & Siemers, N. O. *Methods in Molecular Biology*, 563:99 (2009)

## ACKNOWLEDGEMENTS

This research was supported by NIH Grant No. 5P50GM071558