Computational Molecular Biology and Bioinformatics

Disease Modelling

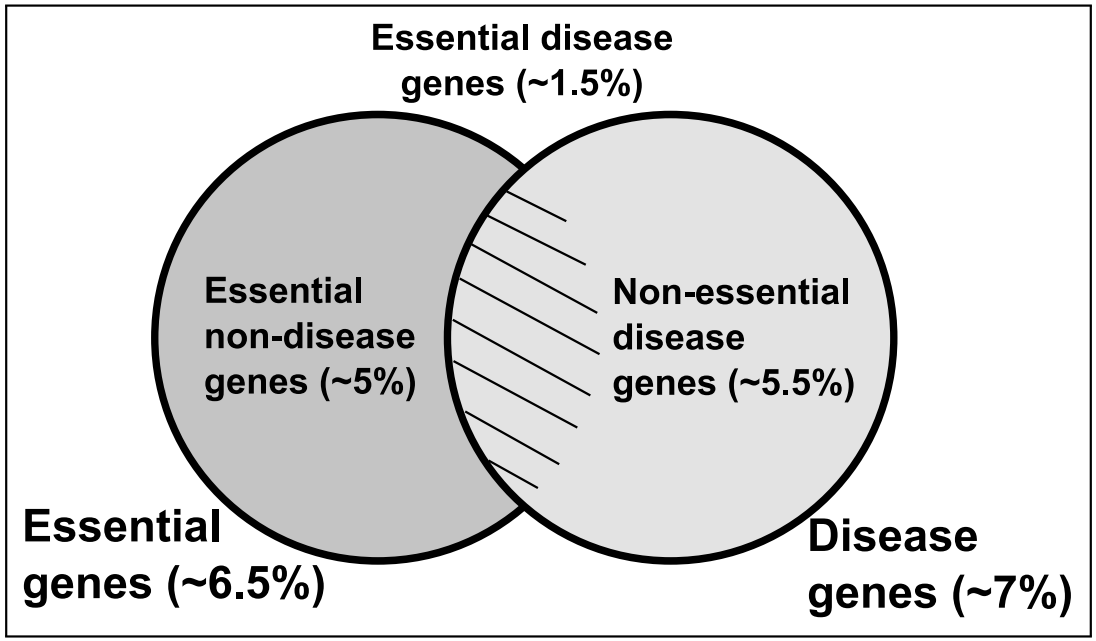
Malay Bhattacharyya

Associate Professor

Machine Intelligence Unit
Indian Statistical Institute, Kolkata

October, 2024

Disease genes



A snapshot from 2007

Currently, less than 10% of human genes are known to have association with specific diseases (Barabási et al., *Nat Rev Genet* 12, 2011).

Of these ∼20% are known to be oncogenes (Cancer Genome Project, February 2014).

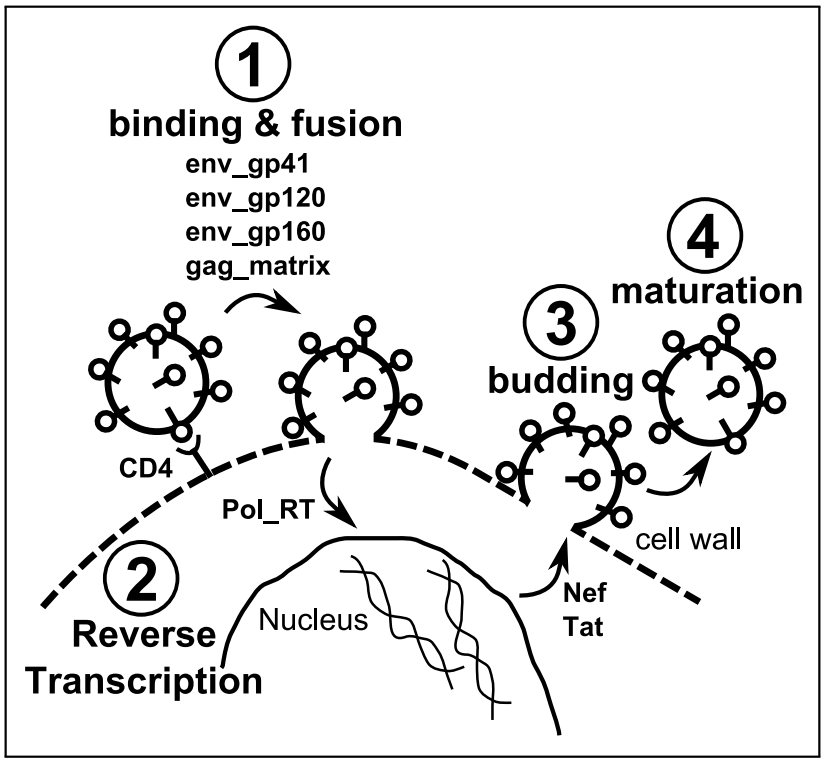
Hypotheses

Some interesting hypotheses applicable to disease networks are as follows.

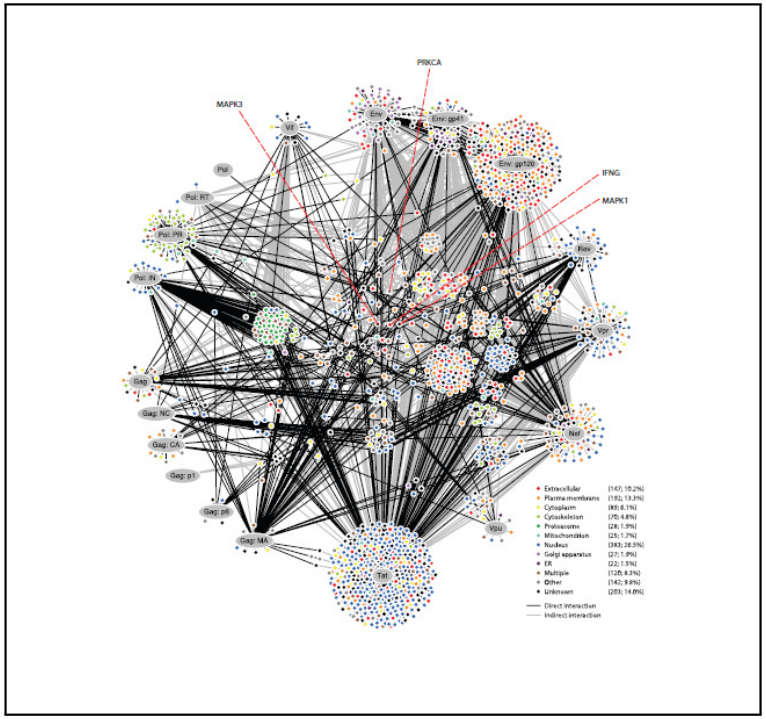
- **Degree:** Disease biomolecules avoid hubs so finding hubs does not point to disease biomarkers.
- **Modularity:** The biomolecules specific to a disease form modules.
- **Sharing:** The diseases having common biomolecules show phenotypic similarity.
- **Closeness:** Causal pathways coincide with the connectors of known disease-specific subnetworks.

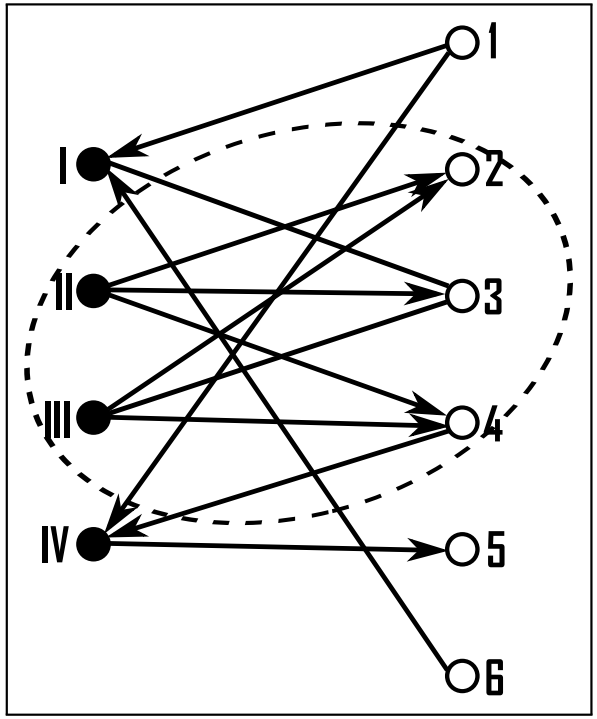
Background of AIDS

"The AIDS is a disease that is hard to talk about" – Bill Gates.



A network view of the viral-host interactions





The HIV-I human protein interaction dataset (Ptak et al., *AIDS Res Hum Retroviruses* 24, 2008).

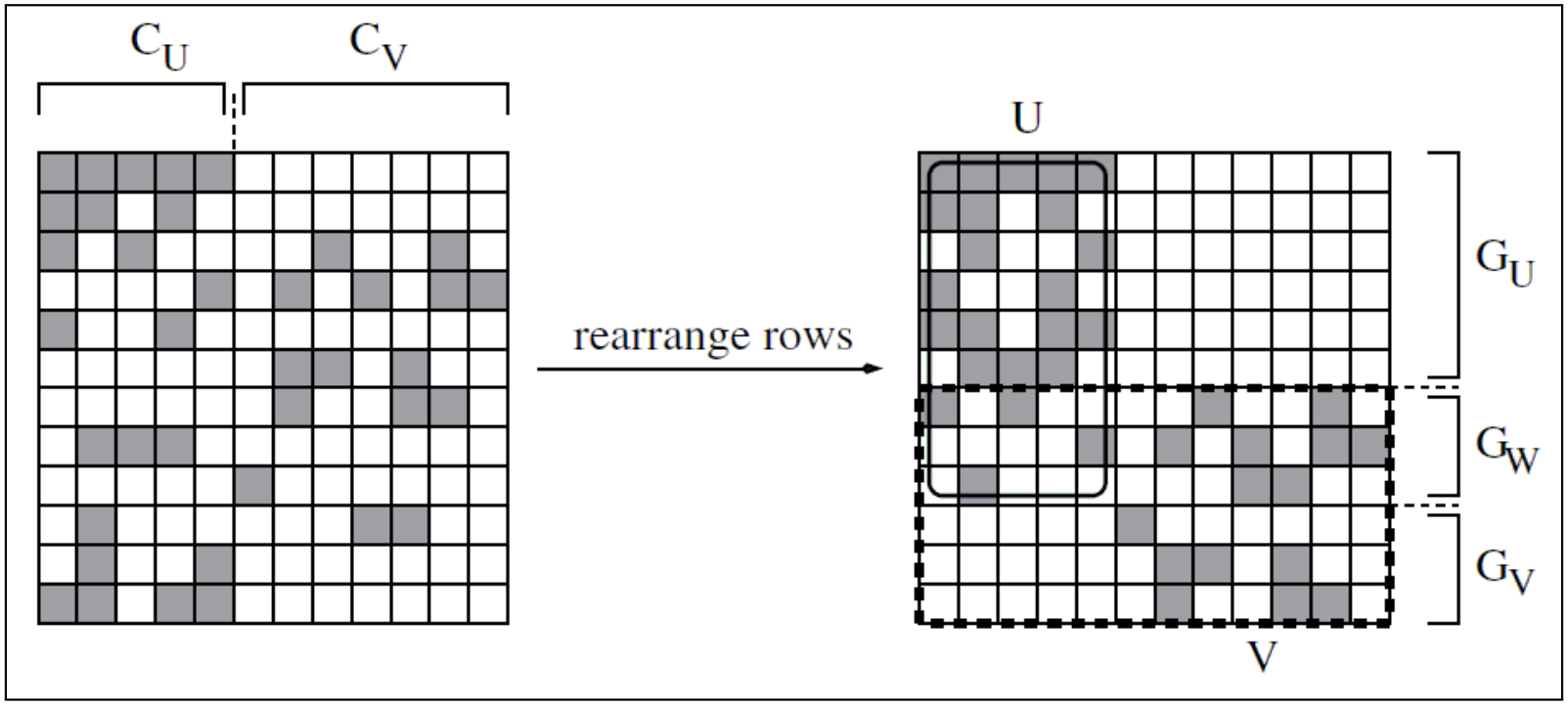
Regulatory network analysis

DBClique

A DBClique is a fully connected subgraph $G' = (V_1', V_2', E') \subseteq G$ of a directed bipartite graph G such that either $i \in V_1', j \in V_2', \forall (i,j) \in E'$ or $i \in V_2', j \in V_1', \forall (i,j) \in E'$.

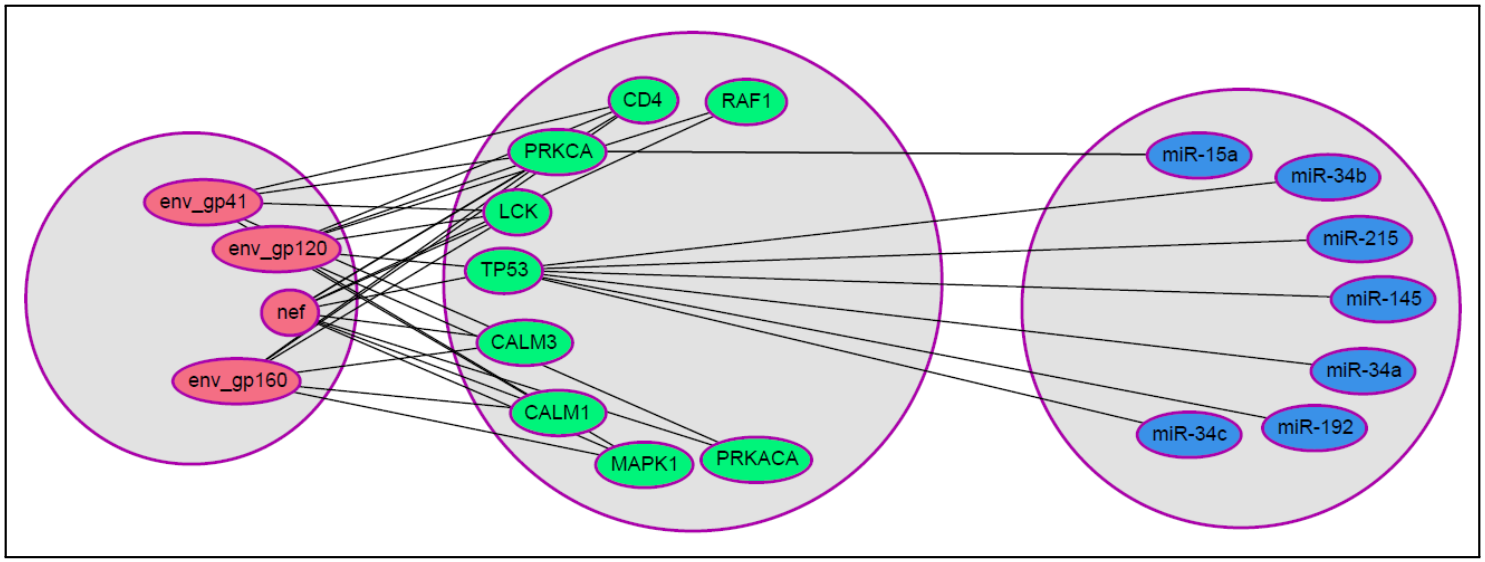
Regulatory network analysis

Directed bicliques can be obtained by rearranging the adjacency matrix of the regulatory network twice to search for ordinary bicliques with some additional constraints.



Disease insights

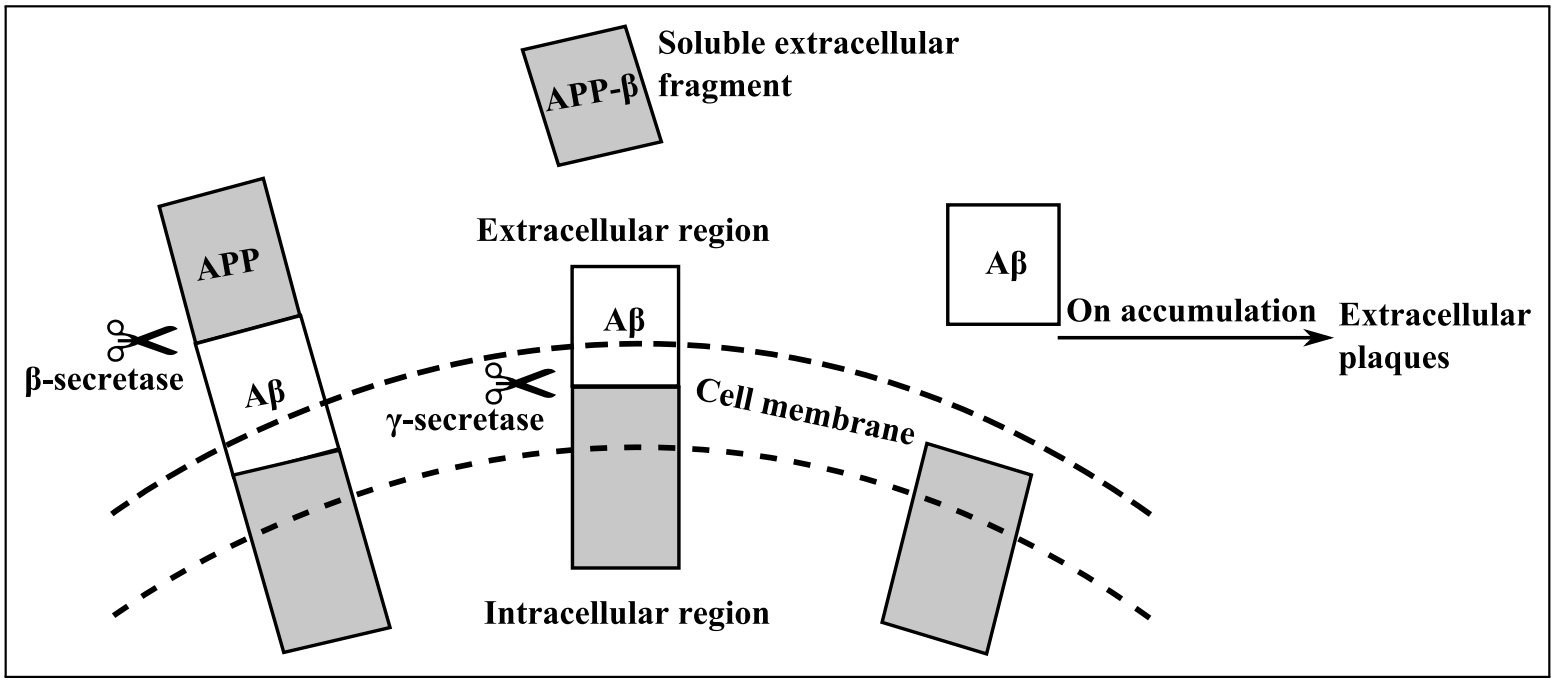
Directed bicliques can be further linked with other regulatory factors to explore the signalling gateway for the host system (Maulik et al., *Mol Biosyst* 7, 2011).



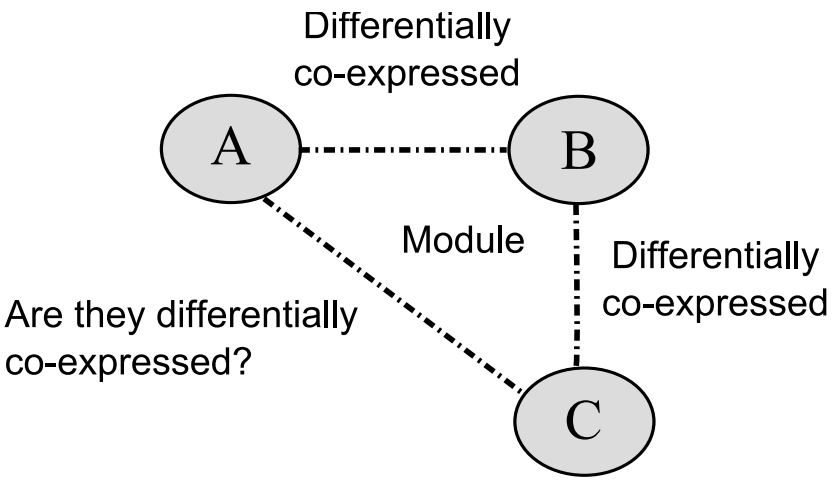
The propagation of HIV-I signal into human cells

Background of alzheimer's disease

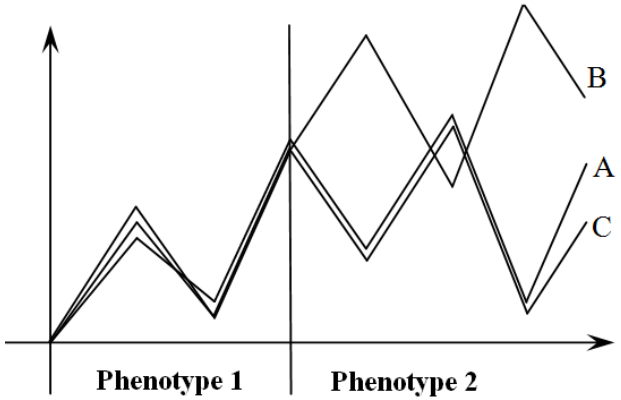
"I haven't heard of anyone who's got better from Alzheimer's"
– Terry Pratchett.



Differential co-expression to module finding



(a)



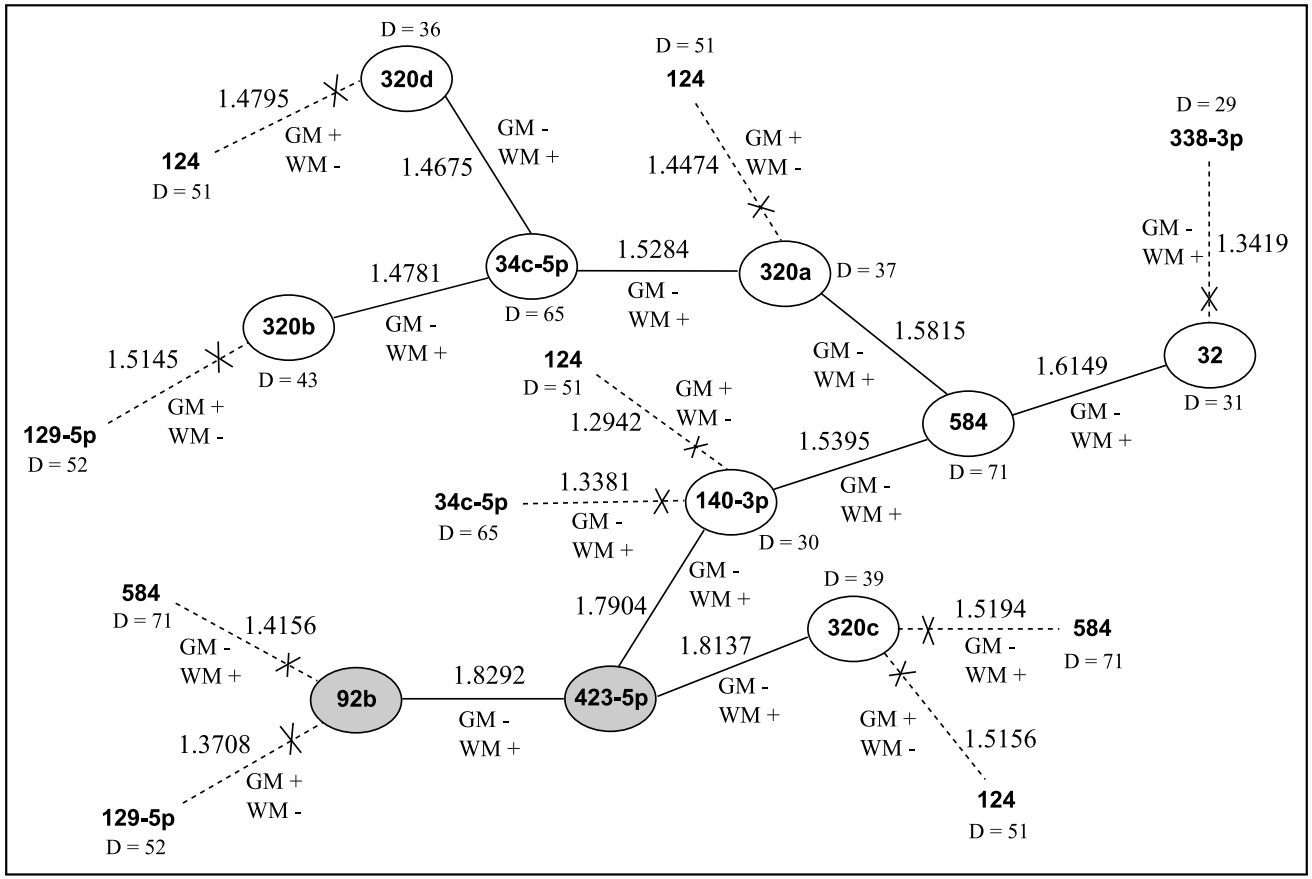
(b)

(a) Possibility of the formation of differentially co-expressed module between the biomolecules A, B and C

(b) Differential co-expression patterns between A, B and C

Differentially co-expressed switching tree

Differentially co-expressed switching tree provides a regulatory overview of the differential pattern (Bhattacharyya et al., *Mol Biosyst* 9, 2013).



Drugs and biomarkers

Drugs and biomarkers can be identified through network analyses.

A test case of cancer metastasis has been proposed pursuing the 'seed and soil' principle (Erler et al., *J Pathol* 220, 2009).

1. Prepare integrative network models.
2. Correlate network dynamics and states to the phenotype and patient disease data.
3. Identify potential multi-node signatures.

Drugs and biomarkers

More focus on aberrations and pathways has also been given recently to find drugs and biomarkers (Pe'er et al., *Cell* 144, 2011).

1. Identify the genetic aberrations and the master regulators that drive proliferation, survival, metastasis, and drug resistance.
2. Model the adaptive/feedback mechanisms that thwart the efficacy of potent drugs.
3. Predict additional target pathways for combinatorial drug treatment.

Hands-on

1. Cancer can be subdivided into various subtypes. Do the following with any specific cancer that has subtypes.
   i) Derive a weighted network between the cancer subtypes by taking data from the TCGA portal (`https://portal.gdc.cancer.gov`).
   ii) Assign weights to the node (cancer subtype) pairs based on the information that how much disease genes they share between each other.
   iii) Now study this network and find whether any interesting motif lies in this.
   iv) Also visualize the entire weighted network with a suitable tool such that the weights are reflected appropriately.