Python to unravel cancer drug target proteins and drug resistance mechanism analysis

Pycon2019, IIT Madras

Subject to change until conference

MD Aksam VK
Data Scientist

Mobile: +91-9042404476

ZettaLabs

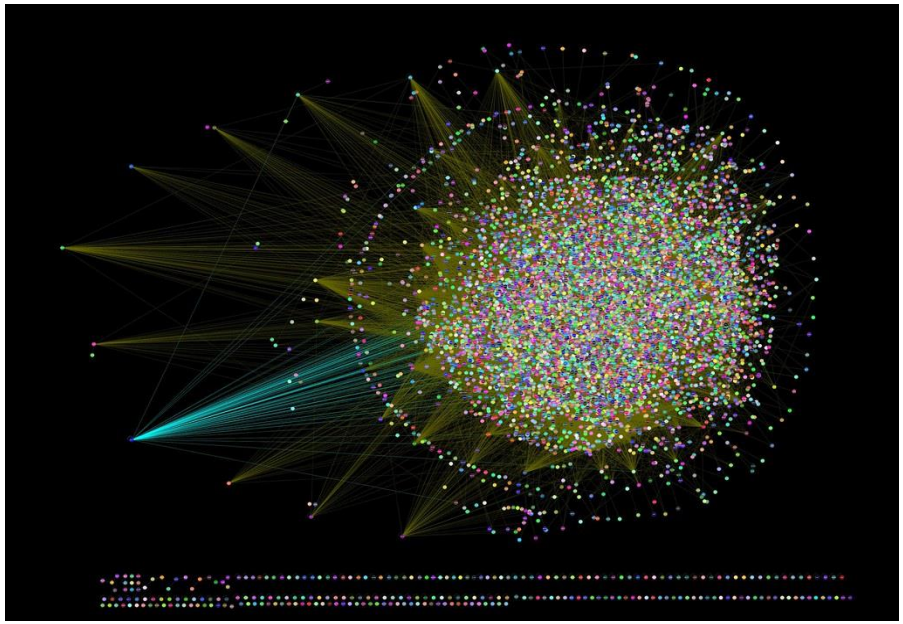
G Floor, Embassy Signet Wing A
Cessna Business Park, Kadubeesanahalli, Outer Ring Road,
Bengaluru, Karnataka 560103
www.zettalabs.com

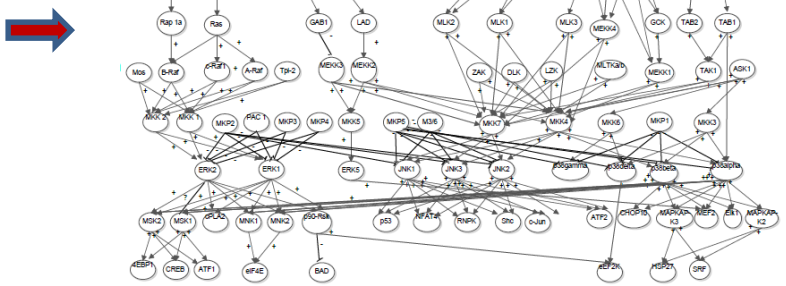
Outline

- Construction of network of MAPK pathways.
- Network pattern revealing with simple metric analysis using python Networkx package.
- Biological process of each protein assigned in the network of MAPK pathways using Matrix.
- Topological and functional attributes of the network based cluster identification.
- Local drug target resistance analysis by calling python function parallelly.

What is network of MAPK pathways..

- Network of MAPK pathways is subnetwork inside a human cell.
- Dysregulation in the network of MAPK pathways causes cancer.
- Disrupting the node(protein) in a network is the therapeutic strategy.
- Furthermore, resistance to the drug is attained due to the concomitant activation of pathways through cross-talks.



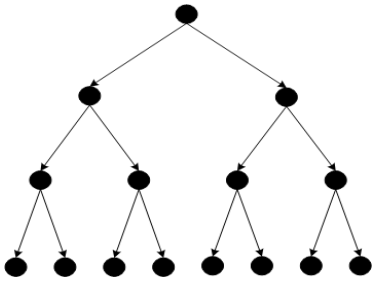


5,400 proteins X 28,500 interactions [HumanInteractome3 - visualised in cytoscape](#) **by** [Andrew Garrow](#)

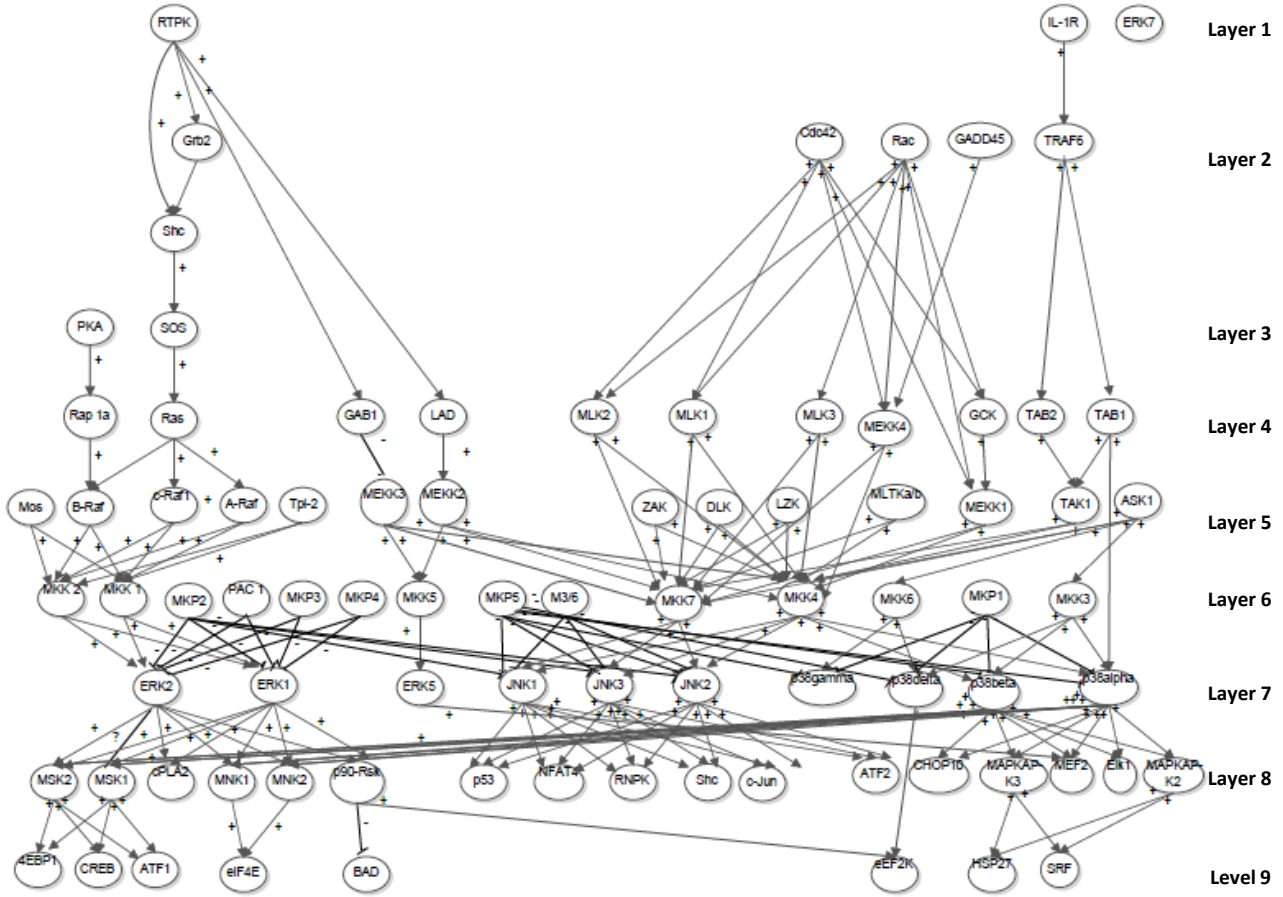
83 proteins X 183 interactions - Network of MAPK pathways [visualised in cytoscape](#) **by MD AKSAM VK**

Network pattern revealing with simple metric analysis using python Networkx package

Hierarchical tree



Layer	Number of nodes	Average in degree	Average out degree
1	1	0	2
2	2	1	2
3	4	1	2
4	8	1	0



layer	No of nodes	Average in degree	Average out degree
1	2	0	2.5
2	6	1.166667	2.833333
3	2	0.5	1
4	11	1.5	1.6
5	14	0.785714	2.428571
6	14	3	3.071429
7	10	4.4	5
8	16	2.875	0.875
9	8	1.875	0

Directed ordered network
with respect to time

Biological process of each protein assigned in matrix

- We classify the biological process into two category

1. Functional process
2. Cancerous contributing process

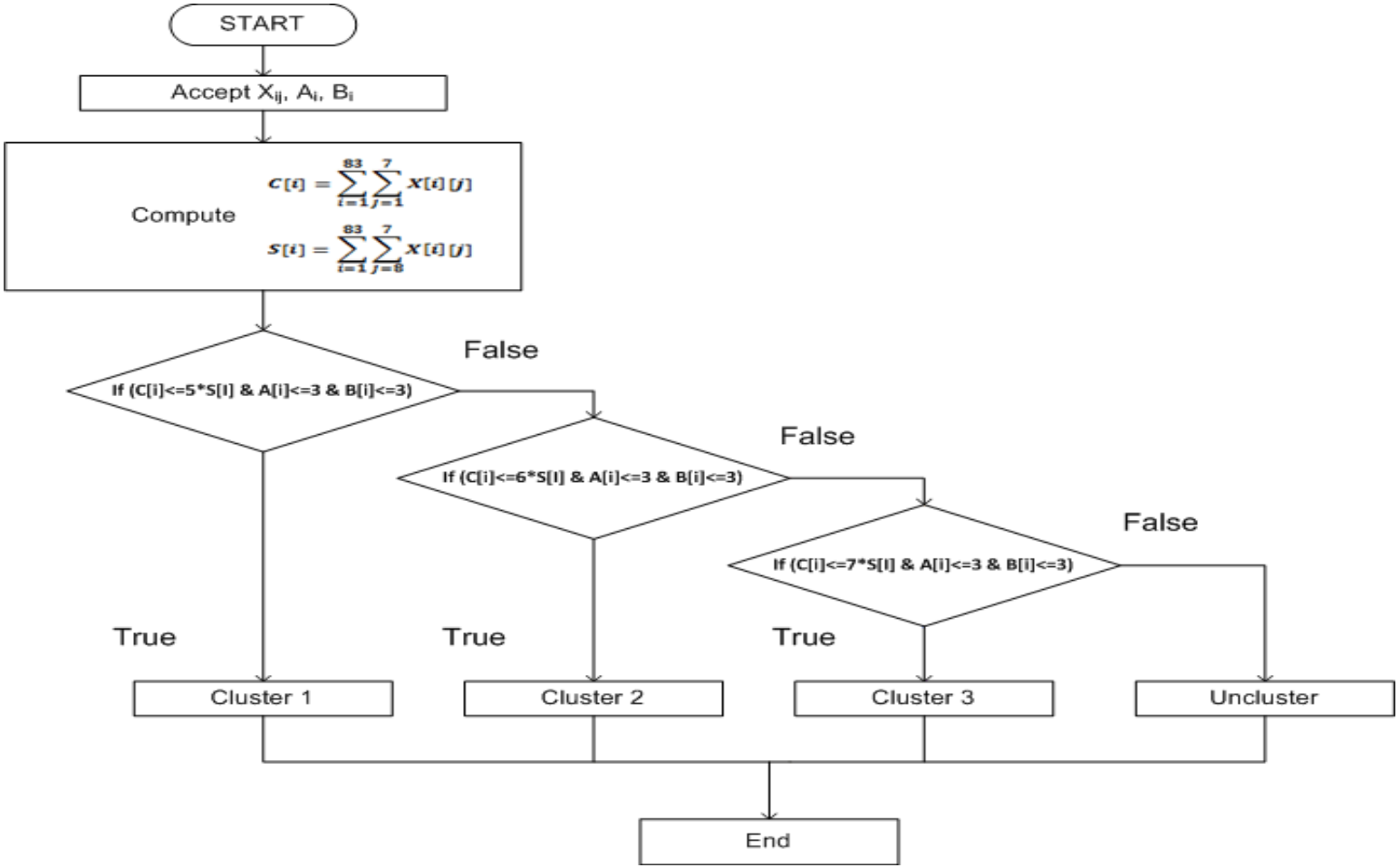
- 83 proteins in the network are given with 82 observed process (Matrix with one hot encoding)
- An adjacency matrix[Aij] is formed with 0 or 1.

$$[\text{Aij}] = \begin{cases} 1\ for\ the\ nodes\ which\ has\ \text{a specific process} \\ 0\ for\ the\ nodes\ \text{doesn't have a specificprocess} \end{cases}$$

List of processes (red colored are cancer related)

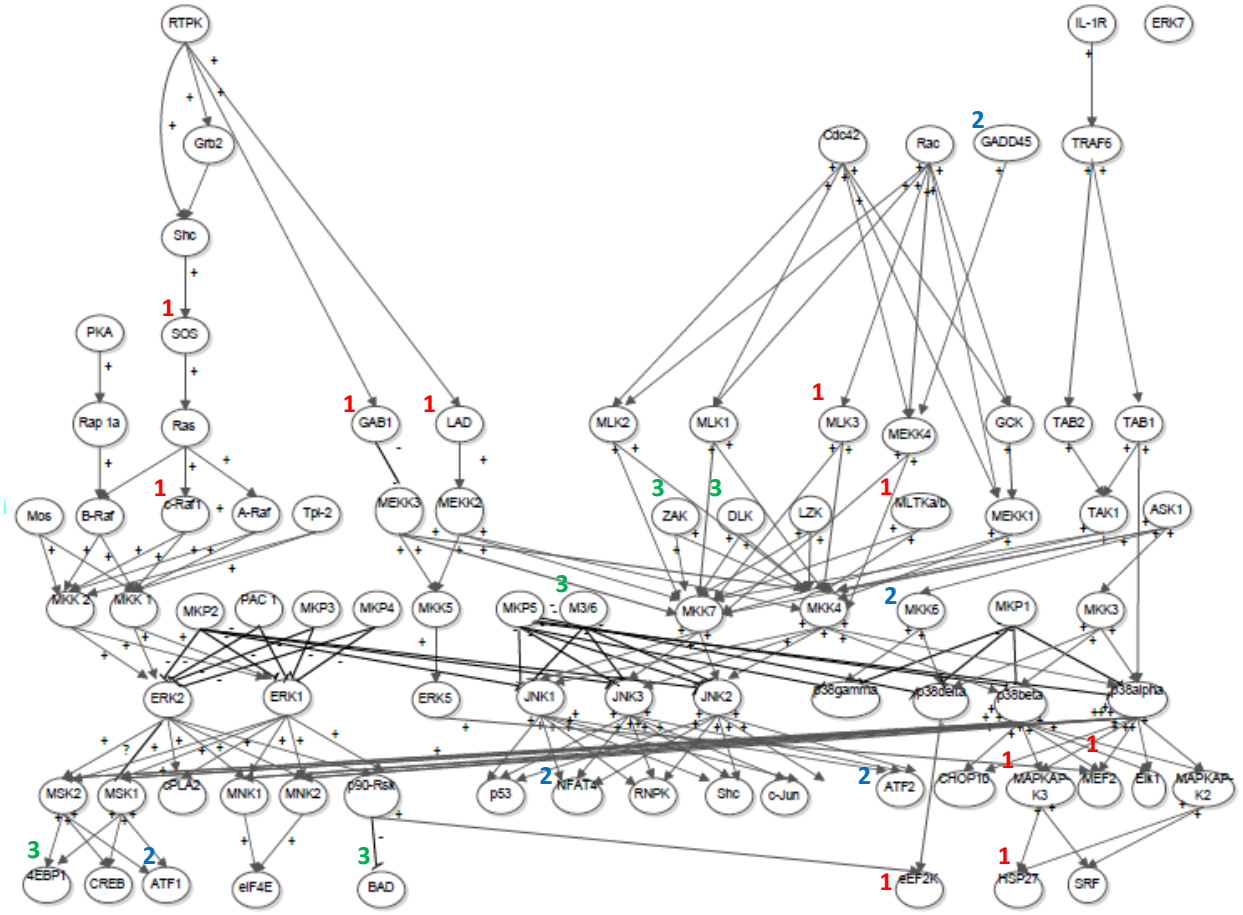
No.	GO:BP	No.	GO:BP	No.	GO:BP	No.	GO:BP
0	response to stress	21	multicellular organism reproduction	42	aging	63	establishment of organelle localization
1	cell proliferation	22	regulation of localization	43	response to endogenous stimulus	64	protein complex biogenesis
2	response to chemical stimulus	23	cellular component morphogenesis	44	cellular component assembly	65	regulation of immune 146
3	regulation of growth	24	macromolecule localization	45	macromolecular complex subunit organization	66	response to biotic stimulus
4	cell death	25	regulation of locomotion	46	interspecies interaction between organisms	67	taxis
5	cell division	26	ovulation cycle	47	cellular response to stimulus	68	establishment of protein localization
6	regulation of anti-apoptosis	27	macromolecule metabolic process	48	protein complex biogenesis	69	response to other organism
7	ossification	28	cellular metabolic process	49	anatomical structure formation involved in morphogenesis	70	sexual reproduction
8	cell activation	29	primary metabolic process	50	embryonic development	71	catabolic process
9	cell cycle	30	positive regulation of biological process	51	transport	72	regulation of homeostatic process,
10	cell adhesion	31	negative regulation of biological process	52	system process	73	regulation of cellular component organization
11	multicellular organismal development	32	positive regulation of cellular process	53	organelle organization	74	translational initiation
12	circadian rhythm	33	negative regulation of cellular process	54	behavior	75	cellular homeostasis,
13	biosynthetic process	34	anatomical structure development	55	actin filament-based process	76	leukocyte activation
14	response to abiotic stimulus	35	cellular developmental process	56	establishment of localization	77	establishment or maintenance of cell polarity
15	anatomical structure morphogenesis	36	regulation of biological process	57	regulation of biological quality	78	cell junction organization
16	positive regulation of metabolic process	37	regulation of cellular process	58	cell motion	79	antigen processing and presentation
17	membrane organization	38	regulation of multicellular organismal process	59	microtubule-based process	80	vesicle targeting
18	regulation of metabolic process	39	cellular localization	60	response to external stimulus	81	regulation of viral reproduction
19	reproductive process	40	regulation of molecular function	61	cellular pigmentation		
20	cell projection organization	41	cell communication	62	developmental process		

Topological and functional attributes of the network based cluster identification



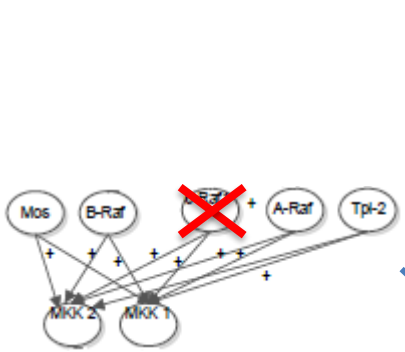
Cont.

Cluster	Node number	Protein name
cluster 1	[4, 9, 30, 32, 33, 38, 58, 64, 78, 81]	SOS, c-Raf1, Eef2k, GAB1, LAD, MEF2, MLK3,MLTKa/b, MAPKAP-K3, HSP27
cluster 2	[28, 46, 56, 68, 76]	ATF1,NFAT4,GADD45,MKK6,ATF2
cluster 3	[26, 31, 40, 49, 50]	4EBP1,BAD,M3/6,ZAK,DLK



Local resistance analysis

The nodes in the clusters are analyzed to see the drug resistance mechanism acquired through the alternative activation of its substrate proteins



Case	Identified cluster node value	Alternatively activated node value	Interpretation
Case 1	1	1	Retain cancer
Case 2	1	0	Efficient drug
Case 3	0	1	Alternate Cancerous process
Case 4	1	1	Retain functions
Case 5	1	0	Loss of function caused side effects
Case 6	0	1	Side effects caused by targeted alternate nodes

Single / combination therapy

11 nodes	Single target
9 nodes	Single / multi target

Switching mechanism over B-RAF to C-RAF are the key observed mechanism, which leads to study the local analysis among alternative switching proteins.[Inamdar et al. 2010]

Protein	Alternative protein	Single/multi drug target
c-Raf1	Mos	single
	A-Raf	single
	B-Raf	single
	Tpl-2	single
M3/6	MKP5	Multi
	MKP2	Single
GADD45	Cdc42	Multi
	Rac	Multi
MLK3	MEKK1	Multi
	MEKK2	Multi
	MEKK3	Single
	MEKK4	Multi
	ASK1	Multi
	MLK1	Multi
	MLK2	Multi
	LZK	Multi
	DLK	Multi
	ZAK	Multi
	MLTKa/b	Multi
	TAK1	Single
MLTKa/b	MEKK1	Multi
	MEKK2	Multi
	MEKK3	Multi
	MEKK4	Multi
	ASK1	Multi
	MLK1	Multi
	MLK2	Multi
	LZK	Multi
	DLK	Multi
	ZAK	Multi
	MLK3	Multi
	TAK1	Multi
MKK6	MKK3	Multi
MAPKAP-K3	MAPKAP-K2	single

Algorithm for local drug resistance analysis

Algorithm: Local analysis

Input: U = Cluster1 U Cluster2 U Cluster3 and Adjacency matrix X[i][j]

Output: Number of cancerous and cellular functions between each node in the cluster and alternate activating nodes sharing are listed

Begin

For each element in U, collect substrate of U activated by set of proteins V

Create function to find similarity of cancerous(1-7) and cellular function role (8-82) between pairs of nodes –

```
func(x,y):
 Let  temp1:=0, temp2:=0, temp3:=0, temp4:=0, temp5:=0, temp6:=0
  for i from 0 to 7
   if x[i]==1 & y[i]==1:
    print 'case1'
    print i
    temp1=temp1+1;
    print temp1
   elif x[i]==1 & y[i]==0:
    print 'case2'
    print i
    temp2=temp2+1;
    print temp2
   elif x[i]==0 & y[i]==1:
    print 'case3'
    print i
    temp3=temp3+1;
    print temp3
```

```
for i from 8 to 82
   if x[i]==1 & y[i]==1:
    print 'case4'
    print i
    temp4=temp4+1;
    print temp4
   elif x[i]==1 & y[i]==0:
    print 'case5'
    print i
    temp5=temp5+1;
    print temp5
   elif x[i]==0 & y[i]==1:
    print 'case6'
    print i
    temp6=temp6+1;
    print temp6
```

Call for each element of U by corresponding alternate activating nodes with function func(U,V) and tabulate output

Single, multi thread and GPU implementation

```
# Drug resistance analysis function
def FUN(x1,c1):
    i=0
    e1=x1[i]
    e2=x1[i+1]
    x2=n1[e1]
    y2=n1[e2];
    temp1,temp2,temp3,temp4,temp5,temp6=0,0,0,0,0,0;
    for i in range(0,7):
        if x2[i]==1 and y2[i]==1:
            temp1=temp1+1;
        elif x2[i]==1 and y2[i]==0:
            temp2=temp2+1;
        elif x2[i]==0 and y2[i]==1:
            temp3=temp3+1;
    for i in range(8,82):
        if x2[i]==1 and y2[i]==1:
            temp4=temp4+1;
        elif x2[i]==1 and y2[i]==0:
            temp5=temp5+1;
        elif x2[i]==0 and y2[i]==1:
            temp6=temp6+1;
    c1[0]=int64(temp1)
    c1[1]=int64(temp2)
    c1[2]=int64(temp3)
    c1[3]=int64(temp4)
    c1[4]=int64(temp5)
    c1[5]=int64(temp6)
```

```
#single thread, multi thread and GPU
gus1 = guvectorize(['void(int64[:], int64[:])'],'(n)->(n)')
gup1 = guvectorize(['void(int64[:], int64[:])'],'(n)->(n)', target="parallel")
guc1 = guvectorize(['void(int64[:], int64[:])'],'(n)->(n)', target="cuda")

gvecs1 = gus1(fun)
gvecp1 = gup1(fun)
gvecc1 = guc1(fun)

with timer("single-thread, scalar y"):
    s0d = gvecs1(inp)

with timer("multi-thread, scalar y"):
    p0d = gvecp1(inp)

with timer("GPU, scalar y"):
    c0d = gvecc1(inp)
```

```
Elapsed time for single-thread, scalar y: 0.0
Elapsed time for multi-thread, scalar y: 0.00
```

Nine nodes c-Raf1, M3/6, GADD45, MLK3, ZAK, DLK, MLTKa/b, MKK6 and MAPKAP-K3 along with their alternate activating proteins

NO.	protein	Alternative protein	case1	no of cancerous attributes	case2	no of cancerous attributes	case3	no of cancerous attributes	case4	no of functional attributes	case5	no of functional attributes	case6	no of functional attributes
1	c-Raf1	Mos	nil	nil	1,4	2	nil	nil	27,28,29	3	36,37,53	3	nil	nil
		A-Raf	nil	nil	1,4	2	nil	nil	27,28,29,36,37	5	53	1	nil	nil
		B-Raf	nil	nil	1,4	2	nil	nil	27,28,29,36,37	5	53	1	11,15,31,33,34,44,45,48	8
		Tpl-2	nil	nil	1,4	2	nil	nil	27,28,29,36,37	5	53	1	9	1
2	M3/6	MKP5	nil	nil	2	1	0	1	27,28,29,36,37	5	18,40	2	47	1
		MKP2	2	1	nil	nil	nil	nil	27,28,29,36,37	5	18,40	2	nil	nil
3	GADD45	Cdc42	0,4	2	nil	nil	2,3	2	9,18,27,28,29,36,40,47,53	10	59	1	11,14,16,19,20,21,22,24,25,30,31,32,33,34,35,38,39,43,51,56,57,60,62,68,70,71,73,76	28
		Rac	nil	nil	0,4	2	6	1	18,36,37,40,53	5	9,27,28,29,47,59	6	11,15,16,20,23,24,30,31,32,34,35,36,39,44,55,56,63,65,73,77,78	20
4	MLK3	MEKK1	0,4	2	2	1	nil	nil	18,27,28,29,31,32,36,37,40,44,45,47,48	13	9,59	2	11,14,16,22,25,31,33,34,53,57,60,71,73,81	14
		MEKK2	0	1	1,4	2	nil	nil	18,27,28,29,36,37,40,47	8	9,30,32,44,45,48,59	7	nil	nil
		MEKK3	nil	nil	0,1,4	3	nil	nil	27,28,29,30,32,36,37	7	9,18,40,44,45,47,48,59	8	nil	nil
		MEKK4	0	1	1,4	2	nil	nil	18,27,28,29,36,37,40,47	8	9,30,32,44.45,48,59	7	nil	nil
		ASK1	0,4	2	1	1	nil	nil	18,27,28,29,30,32,36,37,40,47	10	9,44,45,48,59	5	46	1
		MLK1	0	1	1,4	2	nil	nil	18,27,28,29,36,37,40,47	8	9,30,32,44,45,48,59	7	nil	nil
		MLK2	0	1	1,4	2	nil	nil	18,27,28,29,30,32,36,37,40,47	10	9,44,45,48,59	5	31,33	2
		LZK	0	1	1,4	2	nil	nil	18,27,28,29,36,37,40,47	8	9,30,32,44,45,48,59	7	nil	nil
		DLK	0	1	1,4	2	nil	nil	27,28,29,36,37,47	6	9,18,30,32,40,44,45,48,59	9	nil	nil
		ZAK	0	1	1,4	2	nil	nil	27,28,29,36,37,47	6	18,30,32,40,44,45,48,59	9	53	1
		MLTKa/b	0,1	2	4	1	3	1	9,18,27,28,29,30,32,36,37,40,47	11	44,45,48,59	4	14,35	2
		TAK1	nil	nil	0,1,4	3	nil	nil	18,27,28,29,30,32,36,37,40,47,	10	9,44,45,48,59	5	11,15,31,33,34,38,49,50,65	9

Cont.

NO.	protein	Alternative protein	case1	no of functions	case2	no of functions	case3	no of functions	case4	no of functions	case5	no of functions	case6	no of functions
5	MLTKa/b	MEKK1	0	1	1,3	2	4	1	14,18,27,28,29,30,32,36,37,40,47	11	9,35	2	11,16,22,25,31,33,34,44,45,48,53,57,60,71,73,81	16
		MEKK2	0	1	1,3	2	nil	nil	18,27,28,29,36,37,40,47	8	9,14,30,32,35	5	nil	nil
		MEKK3	nil	nil	0,1,3	3	nil	nil	27,28,29,30,32,36,37	7	9,14,18,35,40,47	6	nil	nil
		MEKK4	0	1	1,3	2	nil	nil	18,27,28,29,36,37,40,47	8	9,14,30,32,35	5	nil	nil
		ASK1	0	1	1,3	2	4	1	18,27,28,29,30,32,36,37,40,47	10	9,14,35	3	46	1
		MLK1	0	1	1,3	2	nil	nil	18,27,28,29,36,37,40,47	8	9,14,30,32,35	5	nil	nil
		MLK2	0	1	1,3	2	nil	nil	18,27,28,29,30,32,36,37,40,47	10	9,14,35	3	31,33	2
		LZK	0	1	1,3	2	nil	nil	18,27,28,29,36,37,40,47	8	9,14,30,32,35	5	nil	nil
		DLK	0	1	1,3	2	nil	nil	27,28,29,36,37,47	6	9,14,18,30,32,35,40	7	53	1
		ZAK	0	1	1,3	2	nil	nil	27,28,29,36,37,47	6	9,14,18,30,32,35,40	7	53	1
		MLK3	0	1	1	1	4	1	9,18,27,28,29,30,32,36,37,40,47	11	14,35	2	44,45,48,59	4
		TAK1	nil	nil	0,1,3	3	nil	nil	18,27,28,29,30,32,36,37,40,47	10	9,14,35	3	11,15,31,33,34,36,38,49,50,65	9
6	ZAK	MEKK1	0	1	nil	nil	4	1	27,28,29,36,37,47,53	7	nil	nil	11,14,16,18,22,25,30,31,32,33,34,36,37,40,44,45,48,57,60,71,73,80	20
		MEKK2	0	1	nil	nil	nil	nil	27,28,29,36,37,47	6	53	1	18,40	2
		MEKK3	nil	nil	0	1	nil	nil	27,28,29,36,37	5	47,53	2	30,32	2
		MEKK4	0	1	nil	nil	nil	nil	27,28,29,36,37,47	6	53	1	18,40	2
		ASK1	0	1	nil	nil	4	1	27,28,29,36,37,47	6	53	1	18,30,32,40,46	5
		MLK1	0	1	nil	nil	nil	nil	27,28,29,36,37,47	6	53	1	18,40	2
		MLK2	0	1	nil	nil	nil	nil	27,28,29,36,37,47	6	53	1	18,30,31,32,33,40	6
		LZK	0	1	nil	nil	nil	nil	27,28,29,36,37,47	6	53	1	18,40	2
		DLK	0	1	nil	nil	nil	nil	27,28,29,36,37,47,53	7	nil	nil	nil	nil
		MLTKa/b	0	1	nil	nil	1,3	2	27,28,29,36,37,47	6	53	1	9,14,18,30,32,35,40	6
		MLK3	0	1	nil	nil	1,4	2	27,28,29,36,37,47	6	53	1	9,18,30,32,40,44,45,48	8
		TAK1	nil	nil	0	1	nil	nil	27,28,29,36,37,47	6	53	1	11,15,18,30,31,32,33,34,38,40,49,50,65	13

Cont.

NO.	protein	Alternative protein	case1	no of functions	case2	no of functions	case3	no of functions	case4	no of functions	case5	no of functions	case6	no of functions
		MEKK1	0	1	nil	nil	4	1	27,28,29,36,37,47,53	7	nil	nil	11,14,16,18,22,25,30,31,32,33,34,40,44,45,48,57,60, 71,73,81	20
		MEKK2	0	1	nil	nil	nil	nil	27,28,29,36,37,47	6	53	1	18,40	2
		MEKK3	0	1	nil	nil	nil	nil	27,28,29,36,37	4	47,53	2	30,32	2
		MEKK4	0	1	nil	nil	nil	nil	27,28,29,36,37,47	6	53	1	18,40	2
7	DLK	ASK1	0	1	nil	nil	4	1	27,28,29,36,37,47	6	53	1	18,30,32,40,46	5
		MLK1	0	1	nil	nil	nil	nil	27,28,29,36,37,47	6	53	1	18,40	2
		MLK2	0	1	nil	nil	nil	nil	27,28,29,36,37,47	6	53	1	18,30,31,32,33,40	6
		LZK	0	1	nil	nil	nil	nil	27,28,29,36,3747	6	53	1	18,40	2
		MLTK a/b	0	1	nil	nil	1,3	2	27,28,29,36,37,47	6	53	1	9,14,18,30,32,35,40	7
		ZAK	0	1	nil	nil	nil	nil	27,28,29,36,37,47,53	7	nil	nil	nil	nil
		MLK3	0	1	nil	nil	1,4	2	27,28,29,36,37,47	6	53	1	9,18,30,32,40,44,45,48,59	9
		TAK1	nil	nil	0	1	nil	nil	27,28,29,36,37,47	6	53	1	11,15,18,30,31,32,33,34,38,40,49,50,65	
8	MKK6	MKK3	0	1	nil	nil	2	1	18.27,28,29,30,32,36,37,40,52	10	16,38,60	3	9,47	2
9	MAPKAP-K3	MAPKAP-K2	nil	nil	0	1	nil	nil	27,28,37	3	29,36	2	18,38	2

Conclusions & Future work

- Python function is built for recursive cancer drug target identification by using topological and functional properties.
- Cause of resistance mechanism prevailing in the network are revealed and analyzed to overcome them.
- Work can be extended to analyze the complete network of signaling pathways regulating cancer mechanism.
- Available open source data can be used to study specific cancer types.
- Fully automated end to end drug resistance analysis can be built.