

# Supplementary Material

## Network clustering: probing biological heterogeneity by sparse graphical models

Sach Mukherjee<sup>1,2</sup> and Steven M. Hill<sup>2,1</sup>

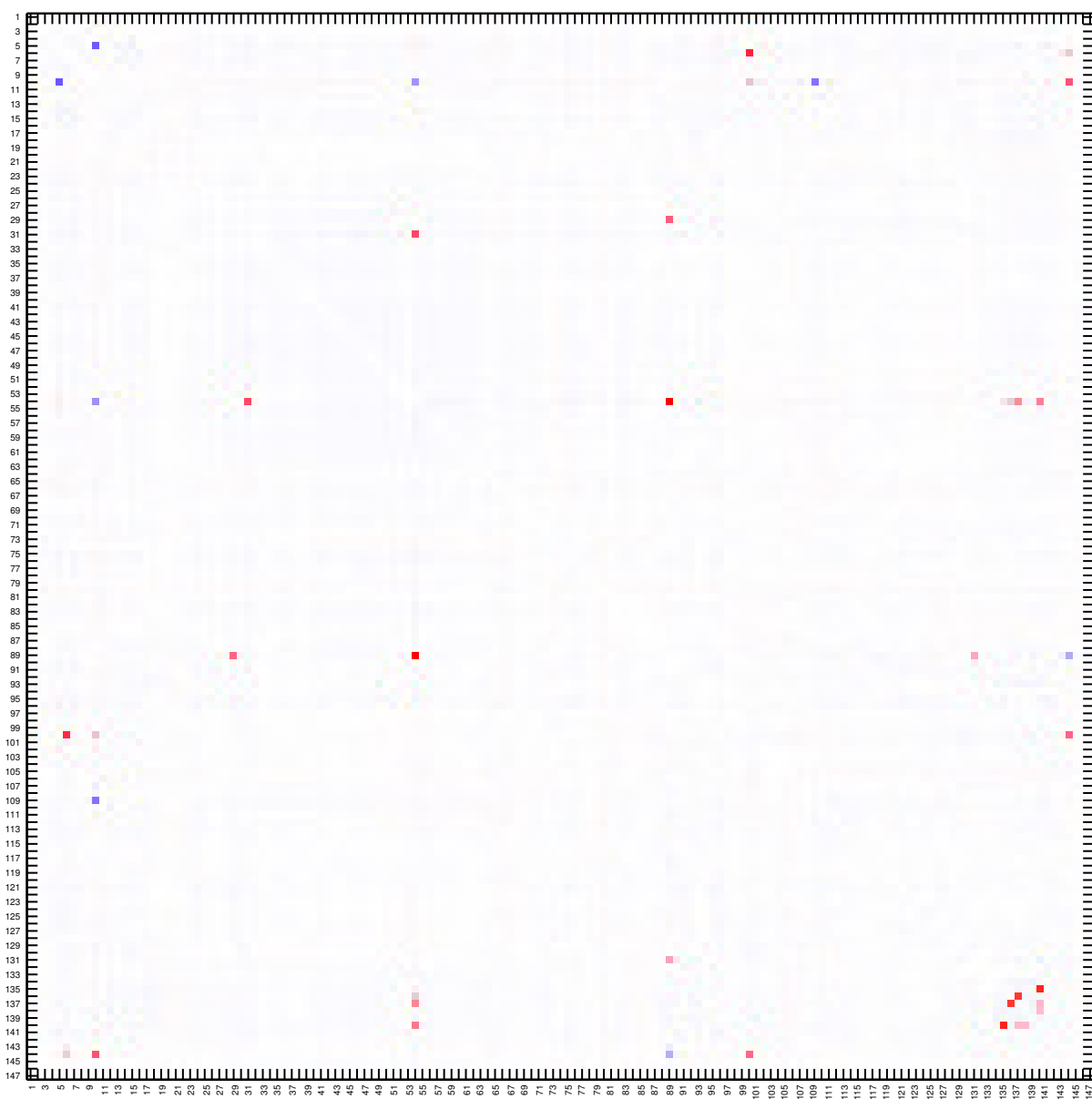
<sup>1</sup>Centre for Complexity Science and <sup>2</sup>Department of Statistics, University of Warwick, Coventry CV4 7AL, U.K.

1	ACVR2A	31	CTTN	61	EP300	91	MCC	121	STAT1
2	ADNP	32	ENAH	62	EP300	92	MCM7	122	STAT1
3	PARP1	33	EP300	63	EP300	93	MCM7	123	STAT1
4	AKAP5	34	EP300	64	ERBB2	94	CD46	124	STAT3
5	AKAP8	35	EP300	65	ESR1	95	MGMT	125	STAT3
6	ANXA1	36	EP300	66	FADD	96	MGMT	126	TRADD
7	ANXA2	37	EP300	67	FADD	97	MLH1	127	TUBB2A
8	AP2M1	38	EP300	68	FADD	98	MSH2	128	TYR
9	ATXN2	39	EP300	69	FN1	99	MSH6	129	VASP
10	BCAR1	40	EP300	70	XRCC6	100	MSN	130	VIL1
11	CASP2	41	EP300	71	GRB2	101	MSN	131	EZR
12	CASP2	42	EP300	72	GSK3B	102	NME1	132	YWHAG
13	CASP7	43	EP300	73	GSTP1	103	PGR	133	STAT3
14	CASP7	44	EP300	74	GSTP1	104	PRKCA	134	STAT5A
15	CASP7	45	EP300	75	GTF2B	105	PRKCB	135	STAT5A
16	CASP7	46	EP300	76	HRAS	106	PRKCI	136	STAT6
17	CCNA2	47	EP300	77	HSPA4	107	PTPN11	137	STAT6
18	CCNB1	48	EP300	78	HSPD1	108	PTPN6	138	TGFB111
19	CDC2	49	EP300	79	IRS1	109	RB1	139	FASLG
20	CDH1	50	EP300	80	IRF9	110	RB1	140	TP53
21	CDH1	51	EP300	81	IRF9	111	RELA	141	TP53
22	CDH1	52	EP300	82	JAK1	112	RELA	142	TP53
23	CDH2	53	EP300	83	KLK3	113	RELA	143	TP53
24	CDH3	54	EP300	84	KRT20	114	RELA	144	TP53
25	CDK4	55	EP300	85	KRT7	115	RELA	145	TP53
26	CDK5	56	EP300	86	KRT8	116	RELA	146	TP53
27	CDK6	57	EP300	87	MAP2K1	117	RIPK1	147	TP53
28	CDK7	58	EP300	88	MAP2K2	118	RNASEH2A		
29	CDKN2A	59	EP300	89	MAP2K2	119	EXOC4		
30	DSG1	60	EP300	90	MAPK1	120	SMARCB1		

Table S1: Proteins from the National Cancer Institute's NCI-60 panel (Shankavaram *et al.*, 2007) used in our cancer proteomic data clustering analysis.

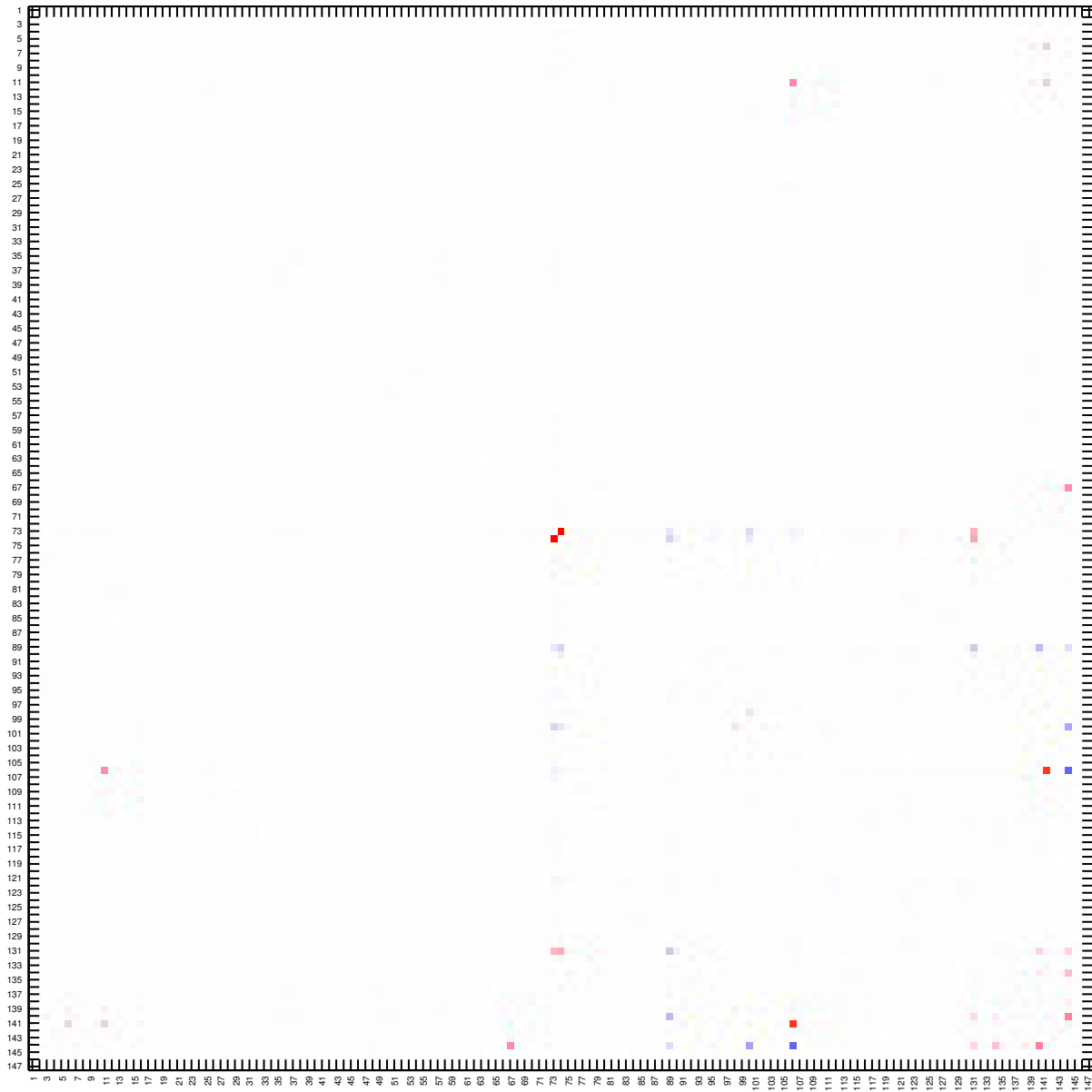
## References

Shankavaram, U. T. *et al.* (2007). Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integrative microarray study. **6**, 820–832.



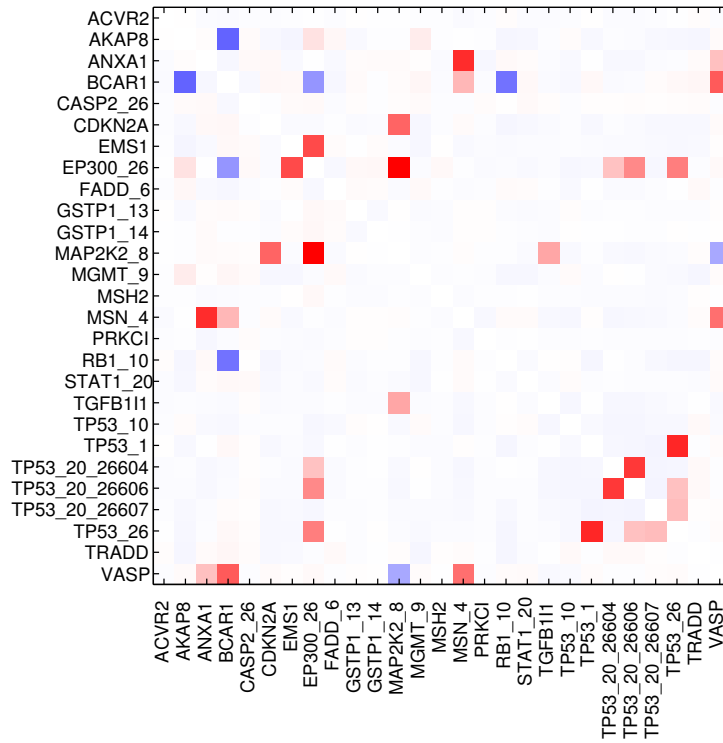
(a) Melanoma cell lines

Figure S1: Network reconstruction results for proteomic data from cancer cell lines. Proteomic data from melanoma and renal cell lines within the NCI60 panel were analyzed by Network Clustering as described in Main Text. Results shown for the highest model score from 100 iterations (this corresponded to the correct clustering into melanoma/renal). We show estimated inverse covariances for (a) Melanoma cell lines and (b) Renal cell lines (red and blue indicate negative and positive values respectively; protein indices as shown in Table S1).

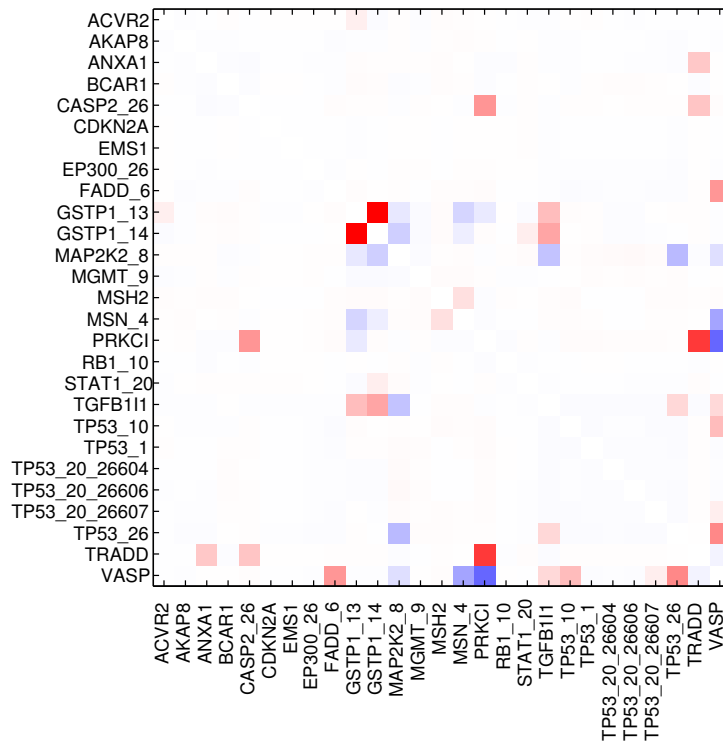


(b) Renal cell lines

Figure S1



(a) Melanoma cell lines



(b) Renal cell lines

Figure S2: Network reconstruction results for proteomic data from cancer cell lines. Inverse covariances as in Figure S1 above, for a subset of proteins only. A protein is included if either or both of the two inverse covariance matrices contain an entry for that protein which is greater than one standard deviation away from the mean of all entries.