

Supplementary files for

Understanding the functional impact of copy number alterations in breast cancers using a network modeling approach

Sriganesh Srihari^{1#}, Murugan Kalimutho^{2#}, Samir Lal³, Jitin Singla⁴, Dhaval Patel⁴,
Peter T. Simpson^{3,5}, Kum Kum Khanna² and Mark A. Ragan^{1*}

¹Institute for Molecular Bioscience, The University of Queensland, St. Lucia,
QLD 4072, Australia

²QIMR-Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia

³The University of Queensland, UQ Centre for Clinical Research, Brisbane,
QLD 4029, Australia

⁴Indian Institute of Technology Roorkee, Roorkee, Uttarakhand 247667, India

⁵The University of Queensland, School of Medicine, Brisbane, QLD 4006, Australia
Supplementary website: <http://bioinformatics.org.au/tools-data/> under NetStrat

Section 4.1 of main text (Materials and Methods)

Choosing ω cut-off

A pair of genes (proteins) could be co-expressed due to a number of reasons – for example, by being co-regulated by the same transcription factor or due to co-functioning in a complex *via* direct interactions. To identify *trans*-associated genes we are only interested in gene pairs (*i.e.* interactions in the network) whose co-expression associates with the CNA of one or both genes. In Equation 2 (main manuscript), since we compute the weighted sum of CNAs, those gene pairs which do not exhibit any CNAs (zero or low CNAs) would be assigned zero or low weights and hence will not be accounted for.

In addition, we would also like to discount gene pairs from the network that show low co-expression values, by using an ω cut-off on the co-expression. This is because these lowly co-expressed interactions often include: (i) interactions that are false-positive or noisy, or (ii) genes whose CNAs do not correlate with the corresponding (co-)expression changes. Here, we choose an ω that is reasonably high, but at the same time allows for sufficient number of interactions. We expect that for most of the gene pairs (A, B) passing this ω cut-off, at least one of A or B exhibits a CNA that agrees with its expression profile (*i.e.* a gain/amplification results in higher expression whereas a loss/deletion results in reduced expression).

At co-expression $\omega = |0.40|$, accounting to 4150 interactions (see full distribution in **Figure 1**), about ~83% (=3443) of all gene pairs (A, B) show (i) CNAs in at least one of A or B ; and (ii) these CNAs correlate with the expression of the genes with $r = 0.69$.

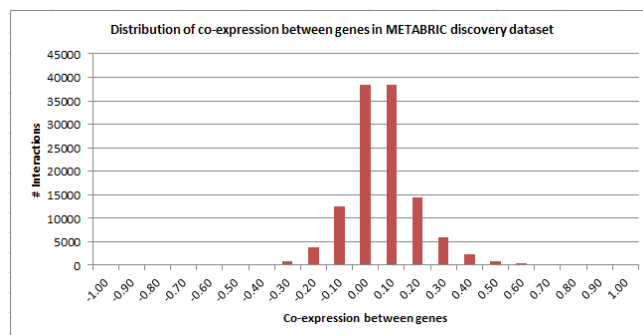


Figure S1: Distribution of co-expression between interacting gene pairs in the PPI network

Section 2.1 of main text (Application to Subtyping of breast tumours)

	<i>Cis</i>	<i>Trans</i>	<i>Cis+Trans</i>
<i>k</i> \ #Genes	917	663	1527
2	0.21	0.13	0.21
3	0.28	0.19	0.28
4	0.38	0.26	0.39
5	0.40	0.31	0.41
6	0.44	0.37	0.46
7	0.49	0.39	0.51
8	0.53	0.43	0.56
9	0.55	0.44	0.59
10	0.58	0.46	0.63
11	0.57	0.39	0.59
12	0.52	0.31	0.53
13	0.48	0.22	0.49
14	0.43	0.14	0.46
15	0.41	0.07	0.43

Table S2: Average Silhouette indices for clusters produced using *cis*-, *trans*- and *cis*- and *trans*-associated genes for $k \in [2, 15]$.

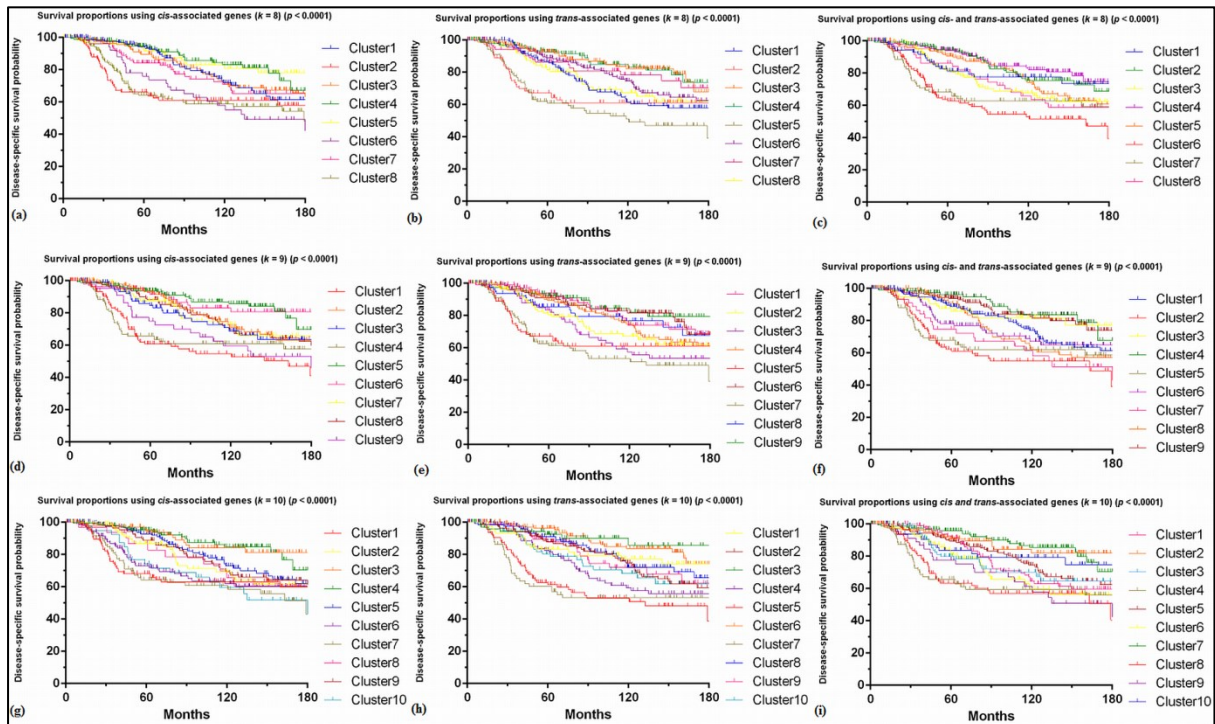


Figure S2: Kaplan-Meier plots of disease-specific survival (truncated at 15 years) for clusters identified using *cis*- (917), *trans*- (663) and by combining *cis*- and *trans*-associated (1527) genes (arranged horizontally) for $k = 8, 9, 10$ clusters (arranged vertically) from the Discovery dataset (998 tumours). Log-rank test p-value in each of the cases was significant ($p < 0.0001$). **Figure 2 of the main manuscript is Figure 2i here.**

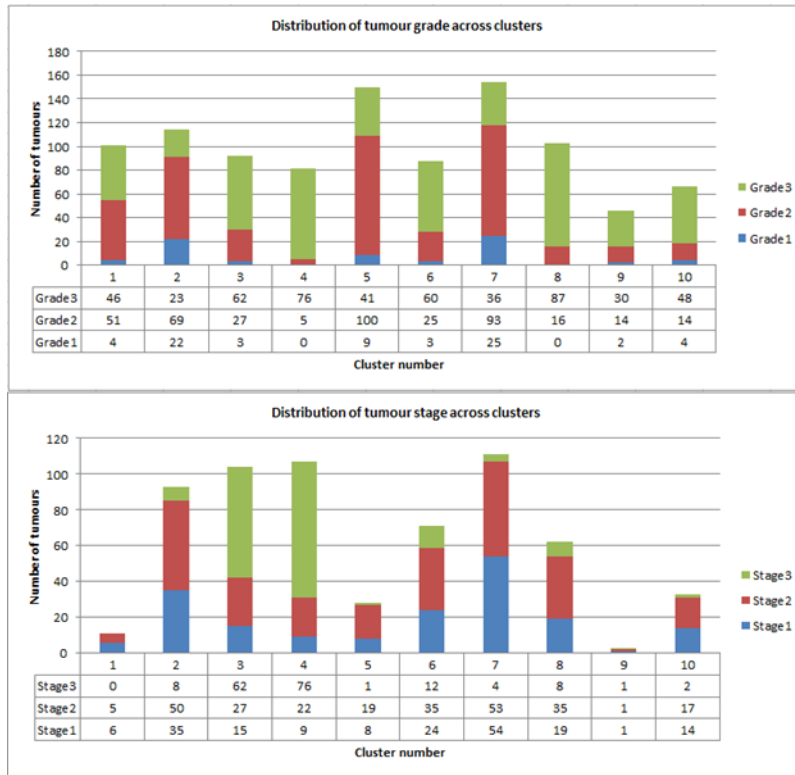


Figure S3: Distribution of tumour grade and stage among the ten clusters

Section 2.2 of main text (In-depth analysis of *cis*- and *trans*-associated genes in the ten clusters)

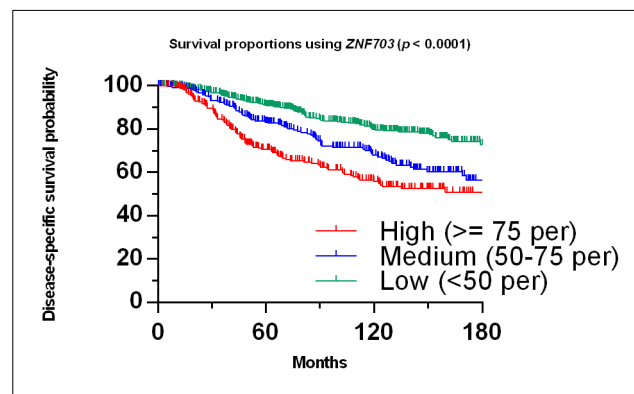


Figure S4: Disease-specific survival proportions of patients stratified based on the expression levels of *ZNF703*. *ZNF703* is over-expressed in clusters 1 and 3, which correspond to luminal tumours that show poor prognosis at 10 years' follow up. Interestingly, very high expression of *ZNF703* (≥ 75 percentile) is seen in 94% (235 out of 250) of the tumours in these clusters.

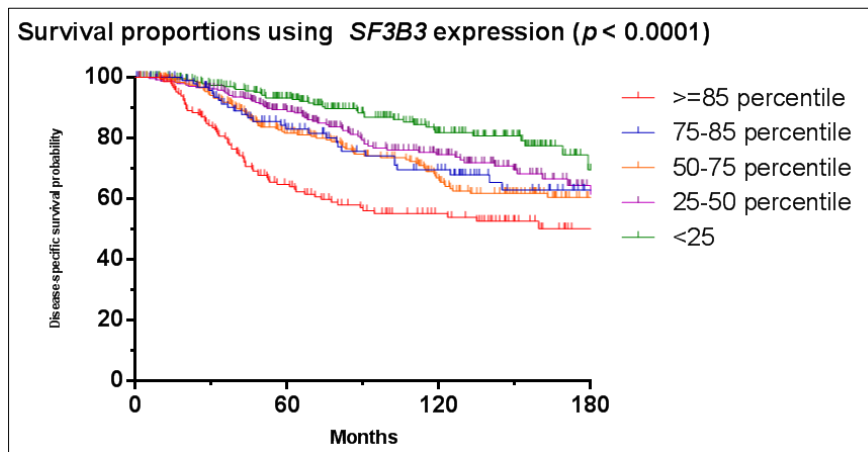


Figure S5: Disease-specific survival proportions of patients stratified based on the expression levels of *SF3B3*. Sixty out of 150 (~40%) tumours showing very high expression (≥ 85 percentile) for *SF3B3* came from cluster 4, which is predominantly composed of basal-like tumours.

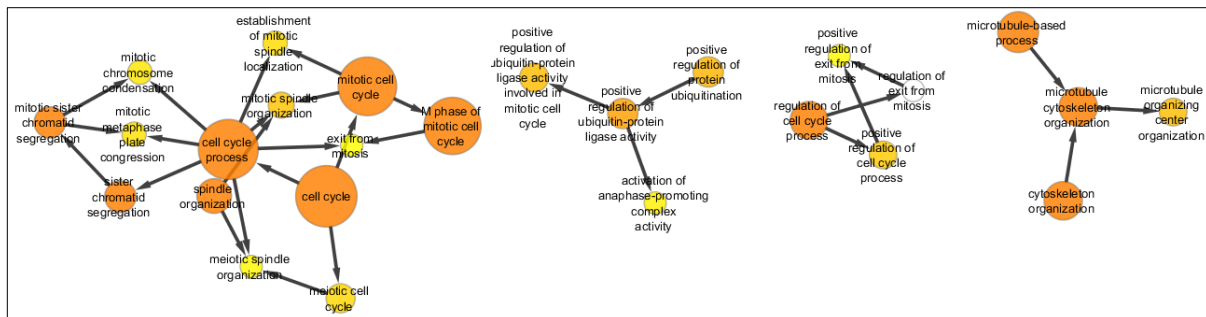


Figure S6: Enrichment of Gene Ontology Biological Process terms in the genes involved in cell division that accumulated high contribution values.

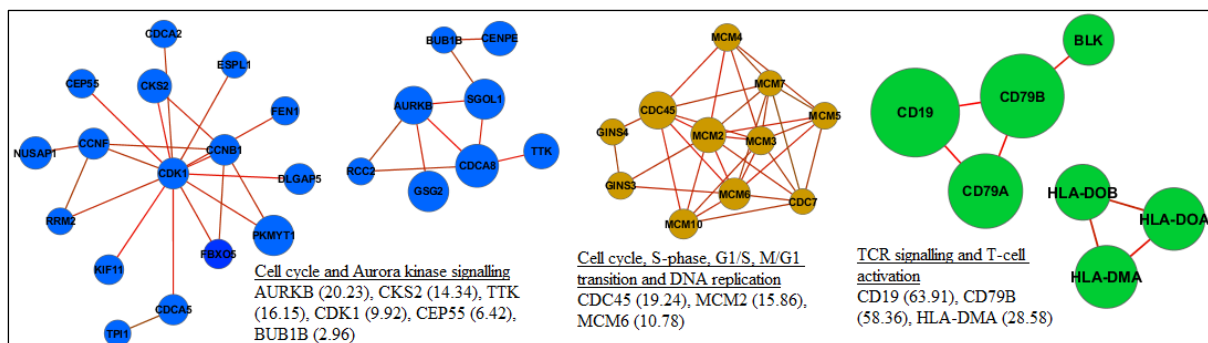


Figure S7: Trans-associated modules of mitotic and immune-response genes

Dataset	Gene	Description	Correlation score (CNA and expression)
COSMIC	<i>AXIN1</i>	Axis Inhibition Protein 1	0.443
	<i>CBFB</i>	Core-Binding Factor, Beta Subunit	0.535
	<i>CCND1</i>	Cyclin D1	0.487
	<i>CDK4</i>	Cyclin-Dependent Kinase 4	0.447
	<i>CLP1</i>	Cleavage And Polyadenylation Factor I Subunit 1	0.519
	<i>DDX10</i>	DEAD (Asp-Glu-Ala-Asp) Box Polypeptide	0.496
	<i>ERCC5</i>	Excision Repair Cross-Complementation Group 5	0.624
	<i>EZR</i>	Ezrin	0.439
	<i>FANCG</i>	Fanconi Anaemia, Complementation Group G	0.461
	<i>FH</i>	Fumarate Hydratase	0.458
	<i>GOLGA5</i>	Golgin A5	0.495
	<i>HERPUD1</i>	Homocysteine-Inducible, Endoplasmic Reticulum Stress-Inducible, Ubiquitin-Like Domain Member 1	0.440
	<i>HOOK3</i>	Hook Microtubule-Tethering Protein 3	0.4727
	<i>IKBKB</i>	Inhibitor Of Kappa Light Polypeptide Gene Enhancer In B-Cells, Kinase Beta	0.549
	<i>KDM5A</i>	Lysine (K)-Specific Demethylase 5A	0.576
	<i>KDM6A</i>	Lysine (K)-Specific Demethylase 6A	0.442
	<i>KRAS</i>	Kirsten Rat Sarcoma Viral Oncogene Homolog	0.528
	<i>MAP2K4</i>	Mitogen-Activated Protein Kinase Kinase 4	0.530
	<i>MDM2</i>	MDM2 Proto-Oncogene, E3 Ubiquitin Protein Ligase	0.497
	<i>MLH1</i>	MutL Homolog 1	0.456
	<i>MLLT10</i>	Myeloid/Lymphoid Or Mixed-Lineage Leukemia (Trithorax Homolog, Drosophila); Translocated To, 10	0.480
	<i>NCOR1</i>	Nuclear Receptor Corepressor 1	0.468
	<i>PALB2</i>	Partner And Localizer Of BRCA2	0.505
	<i>PCMI</i>	Pericentriolar Material 1	0.605
	<i>PPP6C</i>	Protein Phosphatase 6, Catalytic Subunit	0.505
	<i>PRCC</i>	Papillary Renal Cell Carcinoma	0.576
	<i>RAC1</i>	Ras-Related C3 Botulinum Toxin Substrate 1	0.489
	<i>RAD21</i>	RAD21 Homolog (S. Pombe)	0.485
	<i>RAF1</i>	Raf-1 Proto-Oncogene, Serine/Threonine Kinase	0.443
	<i>RECQL4</i>	RecQ Protein-Like 4	0.5249
Vogelstein's list of genes affected by amplification, deletion and rearrangement	<i>CCND1</i>	Cyclin D1	0.487

	<i>MAP2K4</i>	Mitogen-Activated Protein Kinase Kinase 4	0.530
	<i>MDM2</i>	MDM2 Proto-Oncogene, E3 Ubiquitin Protein Ligase	0.497
	<i>PRCC</i>	Papillary Renal Cell Carcinoma	0.576
	<i>RAF1</i>	Raf-1 Proto-Oncogene, Serine/Threonine Kinase	0.443
	<i>SS18</i>	Synovial Sarcoma Translocation, Chromosome 18	0.479

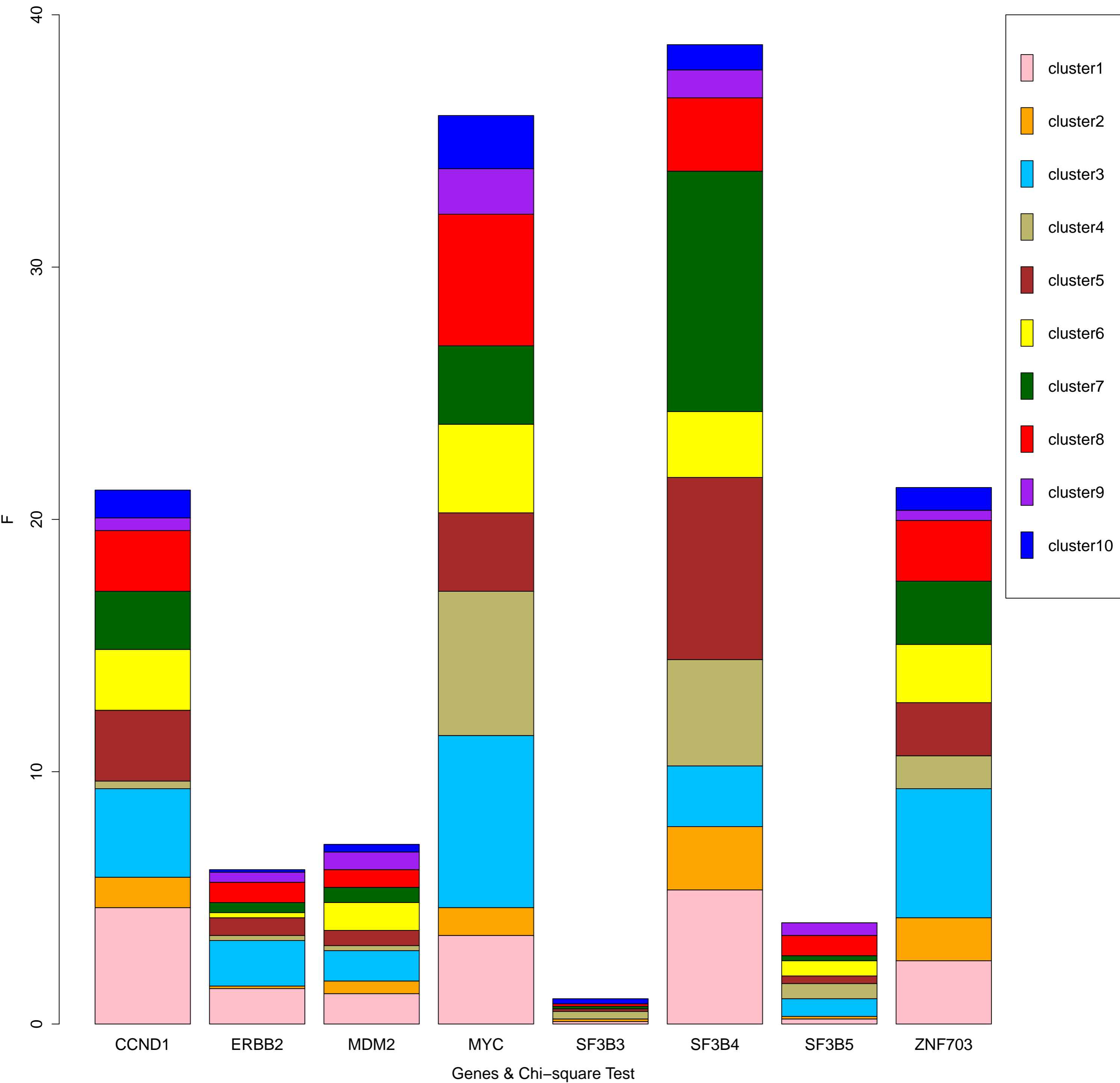
Table S3: List of COSMIC and Vogelstein's genes (affected by amplification, deletion and rearrangements) found among our *cis*-associated genes.

Figure S8: Gains/amplifications observed for *CCND1*, *ERBB2*, *MDM2*, *MYC*, *SF3B3*, *SF3B4*, *SF3B5* and *ZNF703* in the ten clusters.

Figure S9: Losses/deletions observed for *CCND1*, *ERBB2*, *MDM2*, *MYC*, *SF3B3*, *SF3B4*, *SF3B5* and *ZNF703* in the ten clusters.

(S8 and S9 on the next page)

Integrative Clustering Subtypes Segregation –from CCND1 to ZNF703



Integrative Clustering Subtypes Segregation –from CCND1 to ZNF703

