**Frailness and resilience of gene networks predicted by detection of co-occurring mutations via a stochastic perturbative approach**

Matteo Bersanelli1,2,\*, Ettore Mosca3, Luciano Milanesi3, Armando Bazzani1, and Gastone Castellani1

1Department of Physics and Astronomy, University of Bologna, Bologna 40127, Italy

2National Institute for Nuclear Physics (INFN), Bologna 40127, Italy

3Institute of Biomedical Technologies, National Research Council, Segrate (Milan) 20090, Italy

\* matteo.bersanelli2@unibo.it

**Supplementary Information**

**Supplementary Note**

**pME model derivation.** We consider the master equation on an M-nodes network where is the information transition rate between node j and node k. Let be the probability to find the node k is active at time t; the ME reads

(S1)

that corresponds to the continuity equation for the probability distribution with the constraint It is convenient to introduce the vector notation and to define the Laplacian matrix of the network

(S2)

so that the master equation reads as equation (1), and the stationary solution corresponds to the kernel of the matrix *L*. In a generic case *L* has all positive eigenvalues except the the first one which is zero, so that the stationary state is unique and attractive. Different methods can be applied to solve equation (1) without computing the characteristic equation of the matrix L.

We now consider the presence of a network perturbation as an external node (0) playing the role of the environment from which information can be introduced into the network and to which particles can flow out from the current network. Defining the novel connections from () and to () the environment node we can write an extended master equation for the perturbed state in compact form as Equation 2, where is the extension of the Laplacian matrix where the environment node is decoupled from the existing network, while is the perturbation matrix defining the connections with the environment:

where in order to set both and as M+1 x M+1 Laplacian matrices and the vector is now M+1 dimensional as well. If we set , we preserve the stochastic character of the matrix . We omit the subscript in matrix , unlike equation (2) in the main text. Once fixed the configuration of connections to and from the environment, equation (2) has a unique stationary solution that can be computed directly. It is interesting to compute the condition which preserves the stationary solution of the unperturbed system (1). By a direct substitution we have

(S3)

that is equivalent to a detailed balance condition between the source and each sink, so that the probability flow is null for each link.

**Quantifying network frailty.** We look directly at the convergence condition for equation (7) in the main manuscript. We operate a direct substitution

(S4)

where do not depend on the coefficients . So we consider the matrix in the square brackets:

(S5)

the iterative scheme recurrence is driven by and the inverse of matrix in the subspace orthogonal to its kernel so that the condition for equation (7) of the main manuscript to be a contraction becomes:

(S6)

where we are considering a classical matrix norm

when is symmetric its norm corresponds to the absolute value of its biggest eigenvalue.

The right-hand side of equation (S6) is reported in equation (9) of the main text. Thanks to such convergence condition we retrieve the network frailness using the matrix norm . We define the quantity in equation (10) of the main manuscript as the frailness of a network; is the Fiedler number of the unperturbed network (the smallest non-zero eigenvalue of the associated Laplacian matrix).

**Gene Networks.** For the analysis we considered 92 gene networks obtained by integrating KEGG human pathway information (selected as explained in the main text) and Huri protein-protein interactions. The pathways where at least one BRCA patient has at least 2 mutated genes are 81.

Here is the complete list: Purine metabolism, Pyrimidine metabolism, Basal transcription factors, Proteasome, Base excision repair, Nucleotide excision repair, Mismatch repair, Homologous recombination, **MAPK signaling pathway**, ErbB signaling pathway, Calcium signaling pathway, Cytokine-cytokine receptor interaction, **Cell cycle**, p53 signaling pathway, Ubiquitin mediated proteolysis, Autophagy - animal, **mTOR signaling pathway**, **Apoptosis**, Wnt signaling pathway, Notch signaling pathway, TGF-beta signaling pathway, VEGF signaling pathway, Focal adhesion, Adherens junction, Tight junction, Gap junction, Complement and coagulation cascades, Toll-like receptor signaling pathway, Jak-STAT signaling pathway, Natural killer cell mediated cytotoxicity, T cell receptor signaling pathway, B cell receptor signaling pathway, Fc epsilon RI signaling pathway, Leukocyte transendothelial migration, Long-term potentiation, Long-term depression, Regulation of actin cytoskeleton, Insulin signaling pathway, GnRH signaling pathway, Progesterone-mediated oocyte maturation, Melanogenesis, Chemokine signaling pathway, Lysosome, Neurotrophin signaling pathway, Endocytosis, Fc gamma R-mediated phagocytosis, RIG-I-like receptor signaling pathway, NOD-like receptor signaling pathway, Spliceosome, Oocyte meiosis, Peroxisome, Protein processing in endoplasmic reticulum, RNA transport, mRNA surveillance pathway, Fanconi anemia pathway, Serotonergic synapse, NF-kappa B signaling pathway, **PI3K-Akt signaling pathway**, HIF-1 signaling pathway, Hippo signaling pathway, Estrogen signaling pathway, TNF signaling pathway, Prolactin signaling pathway, Ras signaling pathway, **Rap1 signaling pathway**, Adrenergic signaling in cardiomyocytes, FoxO signaling pathway, Thyroid hormone signaling pathway, Platelet activation, Oxytocin signaling pathway, AMPK signaling pathway, cAMP signaling pathway, Signaling pathways regulating pluripotency of stem cells, Sphingolipid signaling pathway, Glucagon signaling pathway, Phospholipase D signaling pathway, Apoptosis - multiple species, Th1 and Th2 cell differentiation, Th17 cell differentiation, IL-17 signaling pathway, Apelin signaling pathway.

In the prevoius list we highlighted the 6 pathways chosen for detailed analysis. The same results of all the 75 remaining gene networks are retrievable from **Supplementary Table ST1, ST3**.

**Supplementary Tables**

**Supplementary Table ST1.** The results of the pME analysis are reported on all available BC patients and all gene networks considered. Columns description (from left to right): **pathway** = name of the KEGG pathway used to identify significant gene network; **patients** = TCGA id of the patients presenting signal on correspondent pathway; **genes** = mutated genes in correspondent BC patient; **rho** and **sigma** correspond respectively to the frailness measure computed as equation (10) and is the fraction of times a gene co-occurs with other genes in the correspondent gene network; **mut** = number of patients presenting the corresponding mutation on the given gene network; **degree** = node degree of mutated gene in the given gene network

**Supplementary Table ST2.** Statistical assestment of marginal mutations impact on frailness: hypergeometric test results. **TFP** = number of top frailness patients considered as target set for hypergeometric test; **PA** = number of pathways available for hypergeometric test. **M (%)** = average percentage (on PA) of profiles with at least one marginal gene (k=1) in TFP. **p < 0.2 (%)** = percentage of PA on which p-value is smaller than 0.2; **p < 0.1 (%)** = percentage of PA on which p-value is smaller than 0.1; **p < 0.05 (%)** = percentage of PA on which p-value is smaller than 0.05.

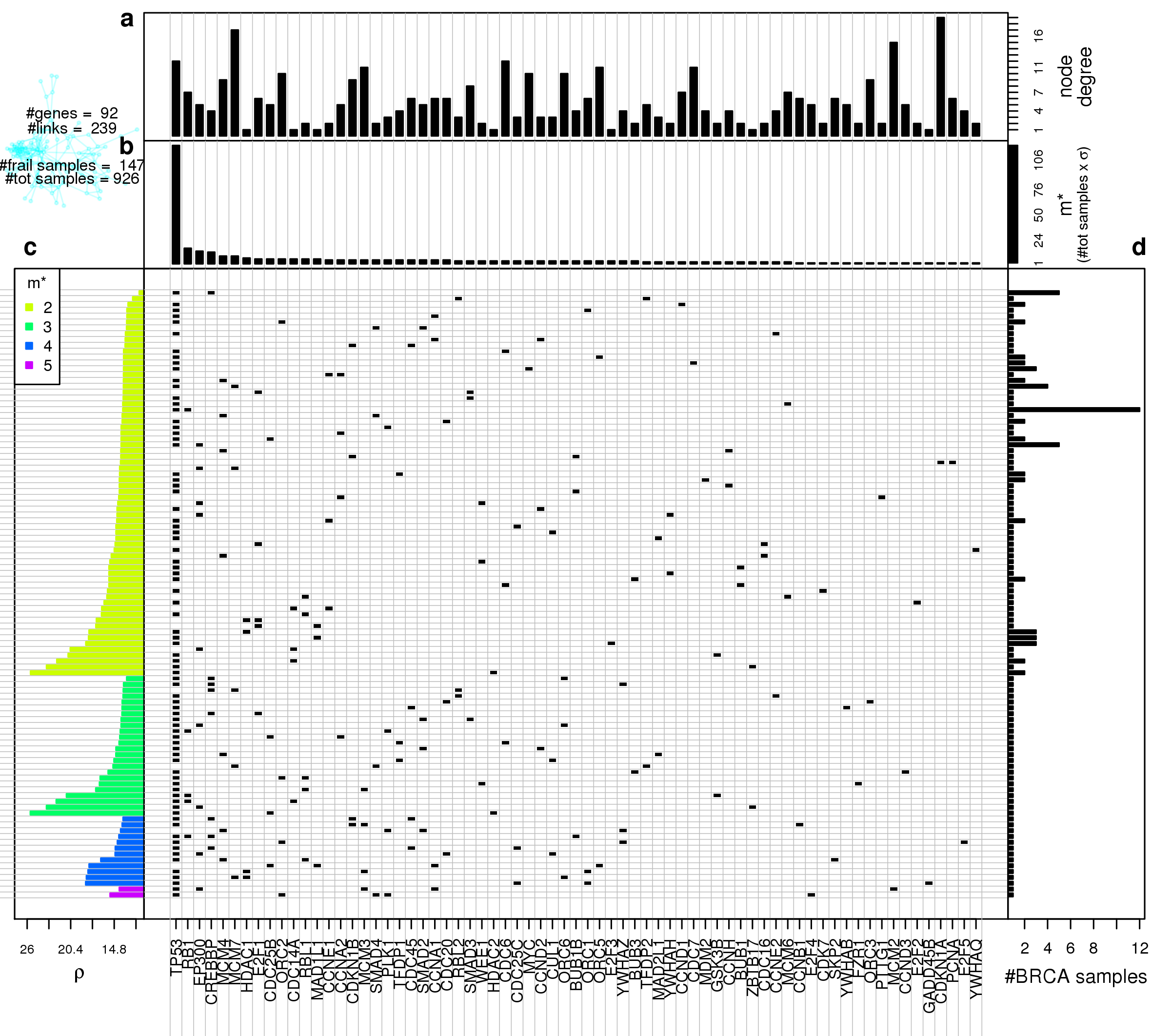
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TFP** | **PA** | **M (%)** | **p < 0.2 (%)** | **p < 0.1 (%)** | **p < 0.05 (%)** |
| 5 | 53 | 93.2% | 98.1% | 98.1% | 96.2% |
| 10 | 42 | 94.3% | 100% | 97.6% | 97.6% |
| 15 | 40 | 91.2% | 100% | 100% | 100% |
| 20 | 32 | 88.1% | 100% | 93.8% | 93.8% |

**Supplementary Table ST3.** The results of the GFI analysis are reported on all available BC patients and all gene networks considered. Columns description (from left to right): **pathway** = name of the KEGG pathway used to identify significant gene network; **genes** = mutated genes in correspondent BC patient; **rho1** and **sigma** correspond respectively to the components and of the GFI as computed in equation (11); **mut** = number of patients presenting the corresponding mutation on the given gene network; **degree** = node degree of mutated gene in the given gene network

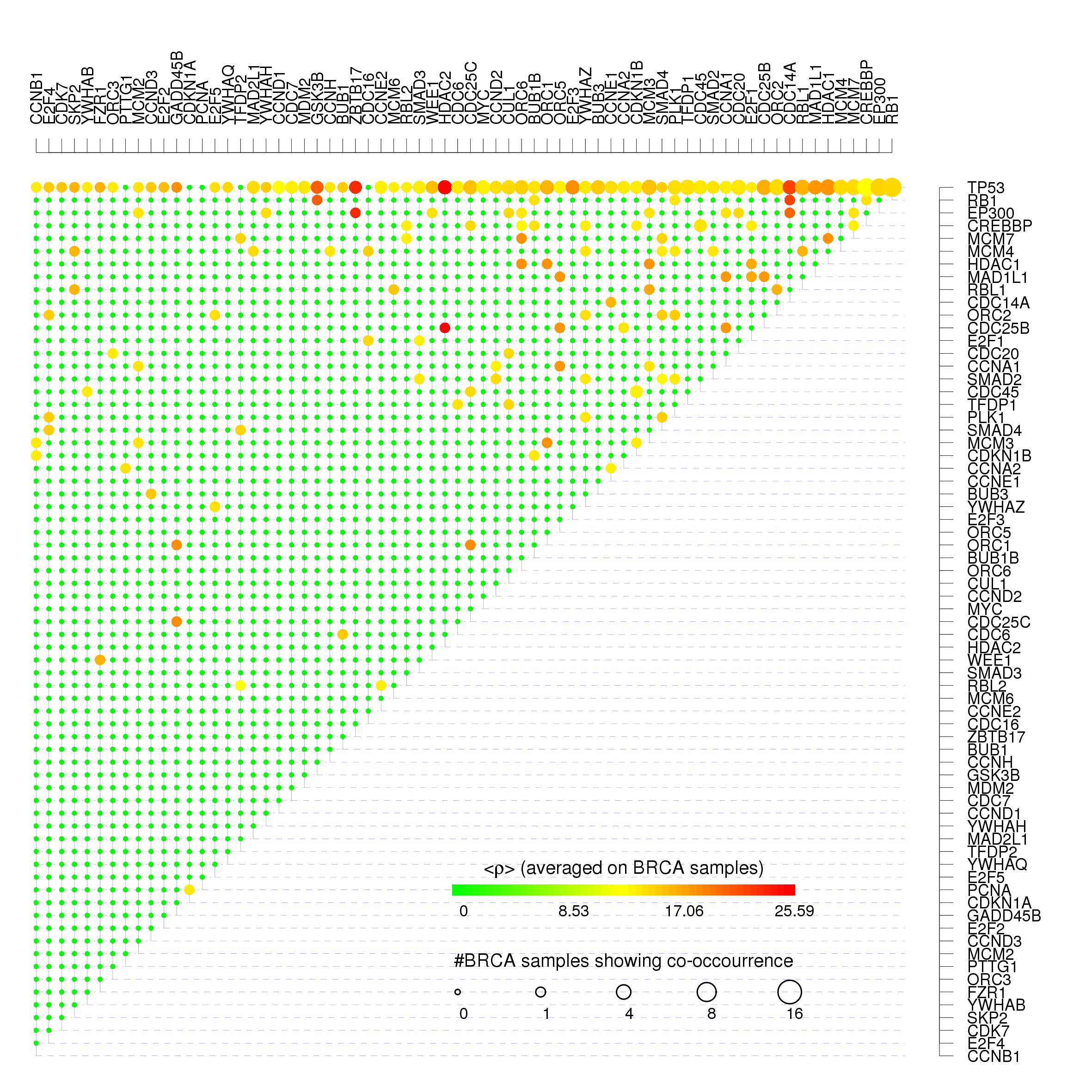
**Supplementary Figures**

**----------------------------------- Cell Cycle ----------------------------------------**

**Supplementary Figure S1**

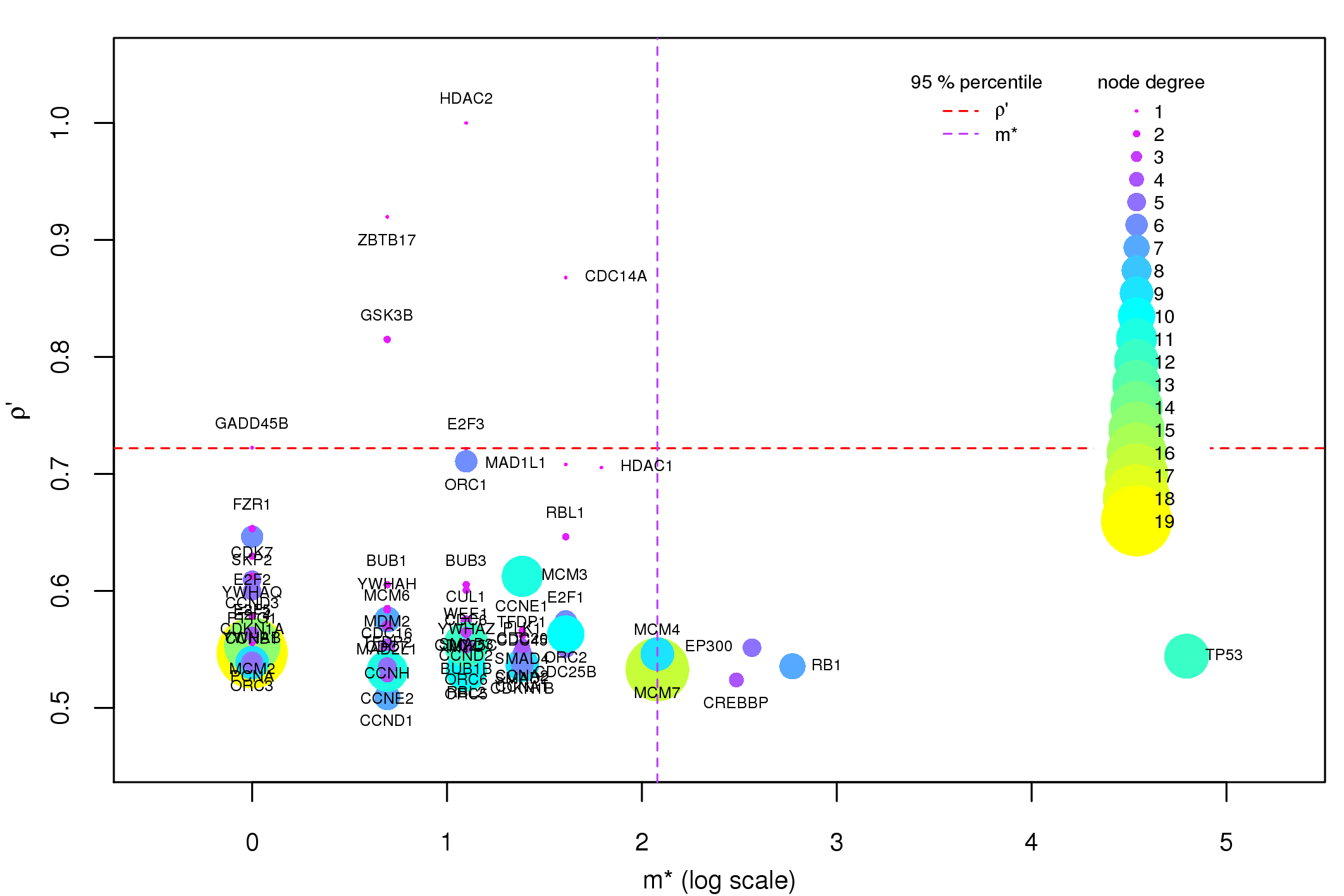
****

**Figure S1. Gene network frailness analysis for Cell Cycle.** Central plot: in each row we plot the somatic mutations co-occurring in the same samples on the Cell Cycle gene network. Rows are ordered as the left-hand side plot, columns are sorted from left to right for decreasing *m*∗. Top panels: from top to bottom is respectively displayed each gene node degree in the target network (a) and the number of patients in which each gene is mutated together with at least another gene (*m*∗) (b); the columns of the central plot are sorted according to decreasing *m*∗. All results were obtained using BC data mapped on KEGG’s Cell Cycle gene network. Left-hand side plot (c): Gene network frailty is plotted for each SM configuration; the ordering on the y-axis is determined in first instance by number of mutations per patient occurring on the gene network, and then in each class the configurations are sorted by ρ. Right-hand side plot (d): for each row of the central plot is displayed the number of patients presenting the corresponding SM configuration.**Supplementary Figure S2**

****

**Figure S2. Gene-gene co-occurrences and frailness measures for Cell Cycle.** For each couple of Cell Cycle co-mutated genes we show the average frailty < ρ > measured on all the BC pathway having the correspondent couple of genes simultaneously mutated using the color scale from green corresponding to null frailty (maximum resilience) to red corresponding to the maximum frailty found in real BC data. The size of each dot is proportional to the log-log scaled number of BC patients presenting the co-mutation.

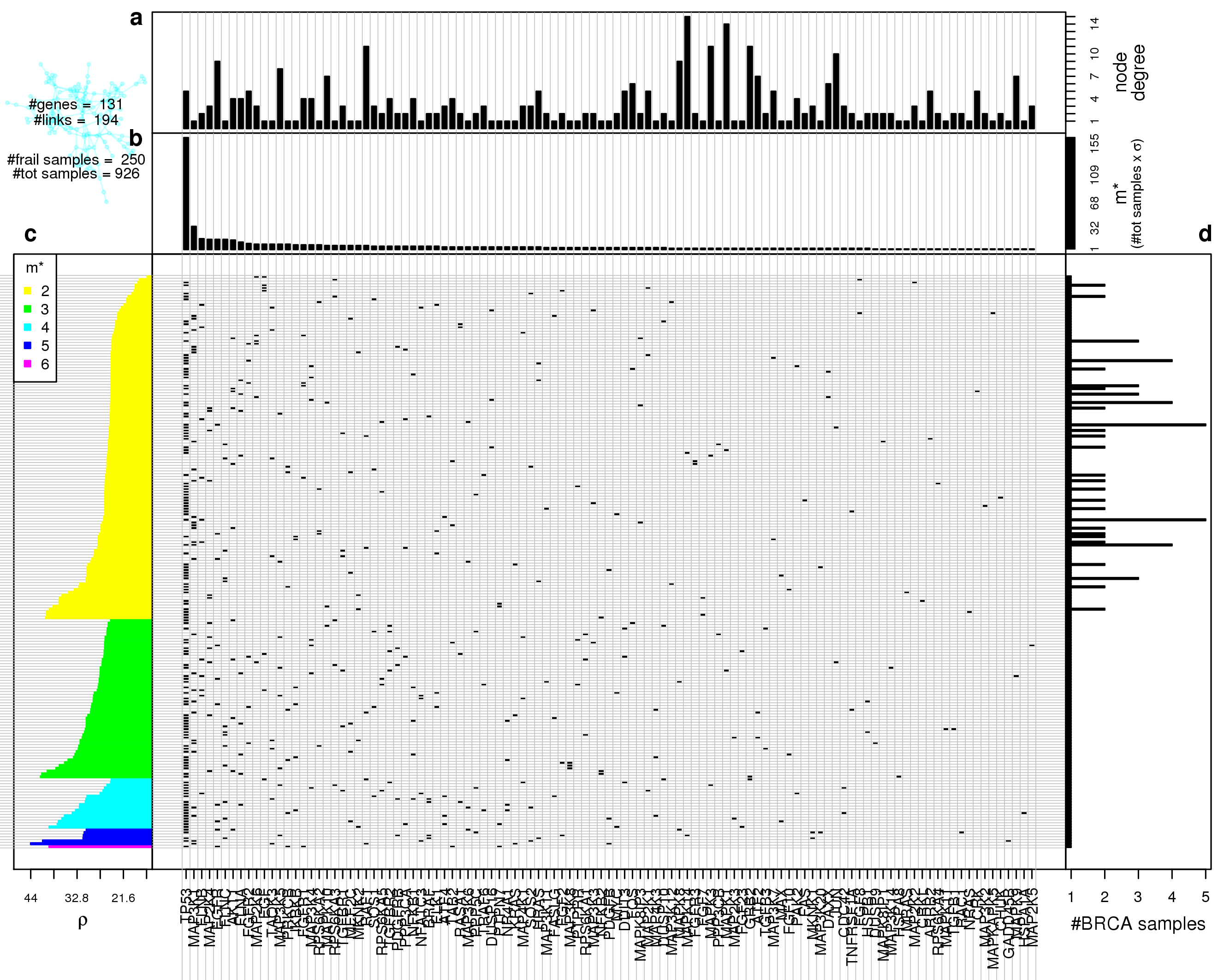
**Supplementary Figure S3**

****

**Figure S3. Components of the GFI for Cell Cycle**. On the Cell Cycle gene network we plot the GFI components for each gene of the network that has at least one non-null component. The x-axis is an elementary transformation of the gene co-occurrence σ (*m*∗ = σ · *nsamples* ) and it is displayed in log scale for visualization issues.

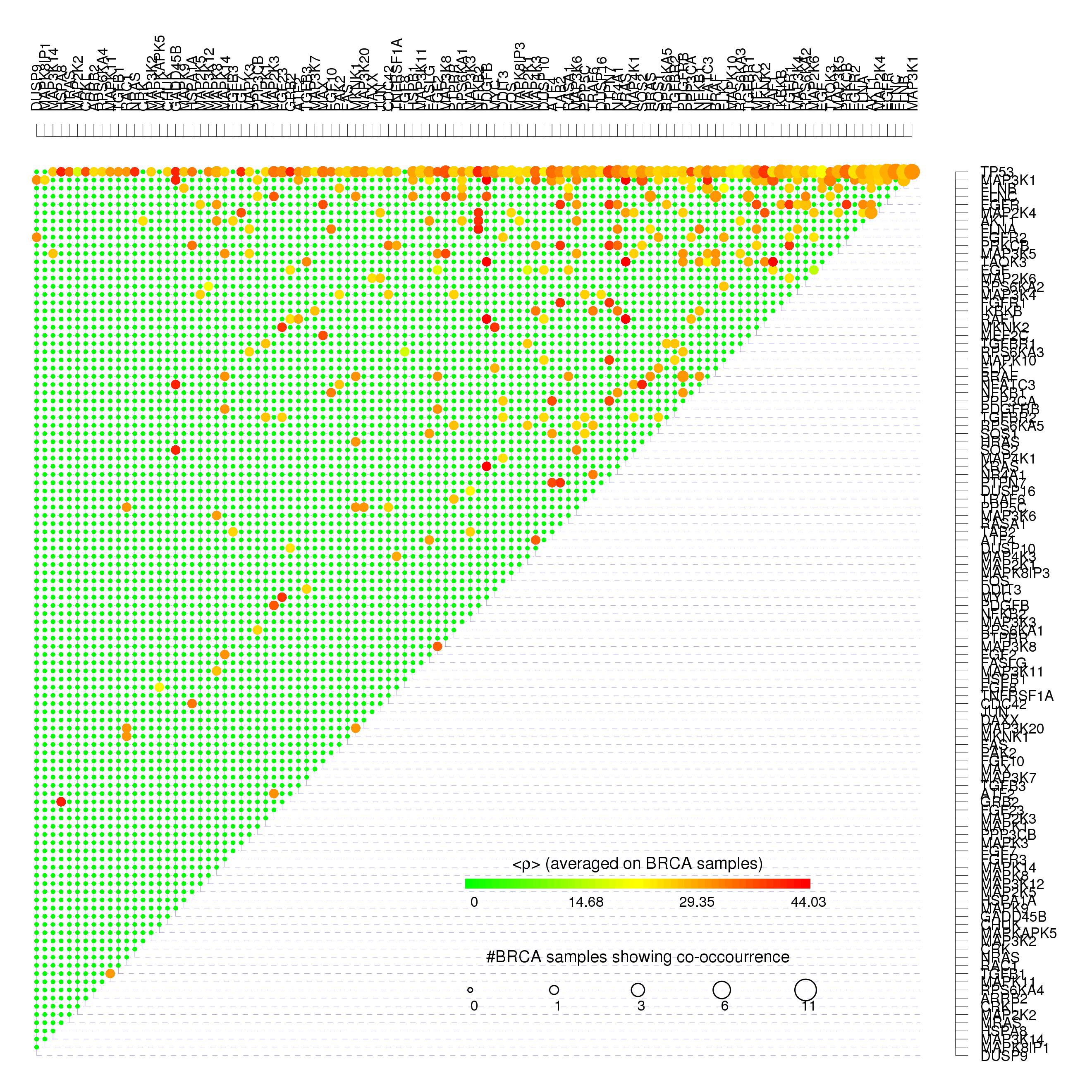
**------------------------- MAPK signaling pathway -----------------------------**

**Supplementary Figure S4**

****

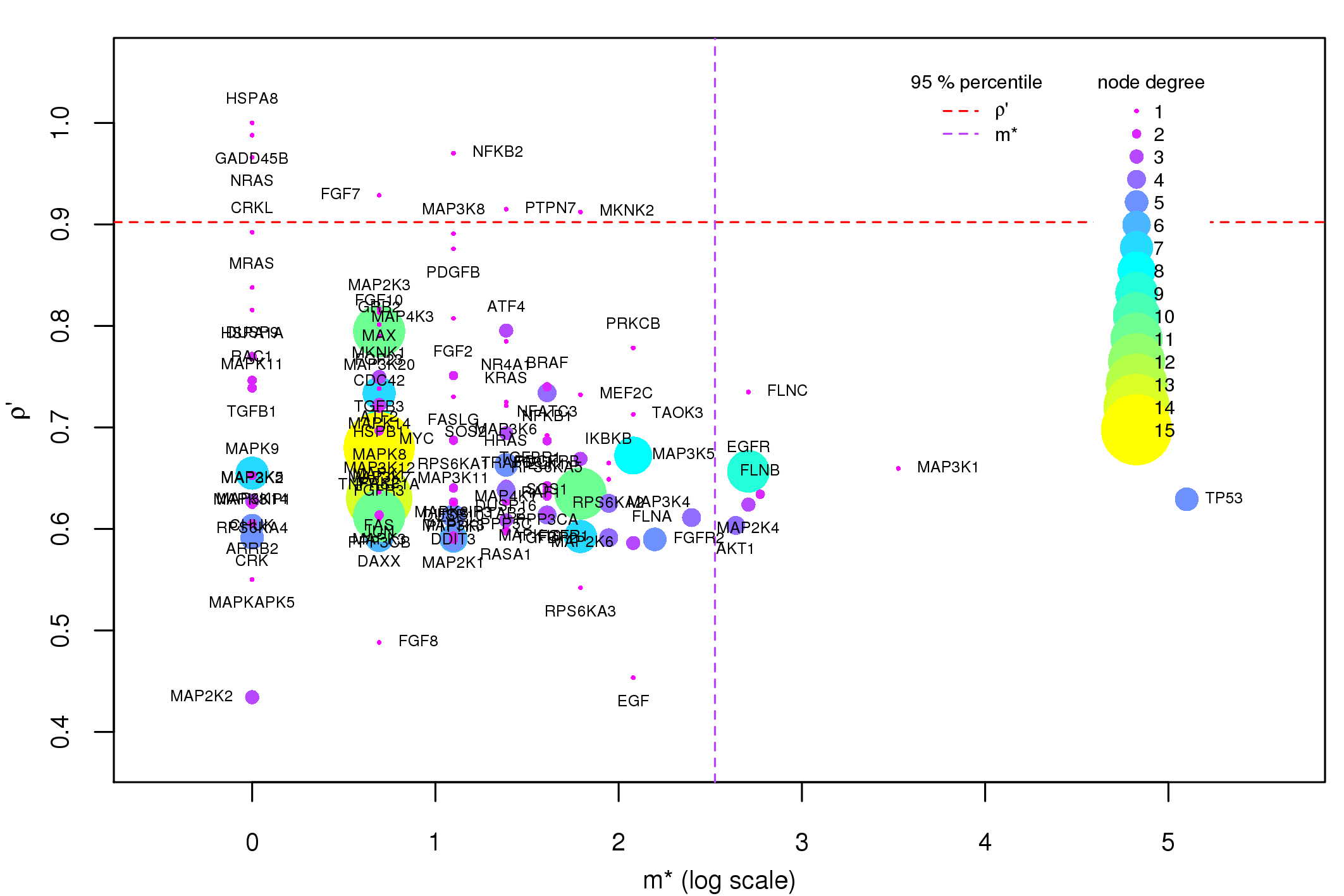
**Figure S4. Gene network frailty analysis for** **MAPK signaling pathway.** Central plot: in each row we plot the somatic mutations co-occurring in the same samples on the MAPK signaling pathwaygene network. Rows are ordered as the left-hand side plot, columns are sorted from left to right for decreasing *m*∗. Top panels: from top to bottom is respectively displayed each gene node degree in the target network (a) and the number of patients in which each gene is mutated together with at least another gene (*m*∗) (b); the columns of the central plot are sorted according to decreasing *m*∗. All results were obtained using BC data mapped on KEGG’s MAPK signaling pathwaygene network. Left-hand side plot (c): Gene network frailty is plotted for each SM configuration; the ordering on the y-axis is determined in first instance by number of mutations per patient occurring on the gene network, and then in each class the configurations are sorted by ρ. Right-hand side plot (d): for each row of the central plot is displayed the number of patients presenting the corresponding SM configuration.

**Supplementary Figure S5**

****

**Figure S5. Gene-gene co-occurrences and frailty measures for** **MAPK signaling pathway.** For each couple of MAPK signaling pathwayco-mutated genes we show the average frailty < ρ > measured on all the BC pathway having the correspondent couple of genes simultaneously mutated using the color scale from green corresponding to null frailty (maximum resilience) to red corresponding to the maximum frailty found in real BC data. The size of each dot is proportional to the log-log scaled number of BC patients presenting the co-mutation.

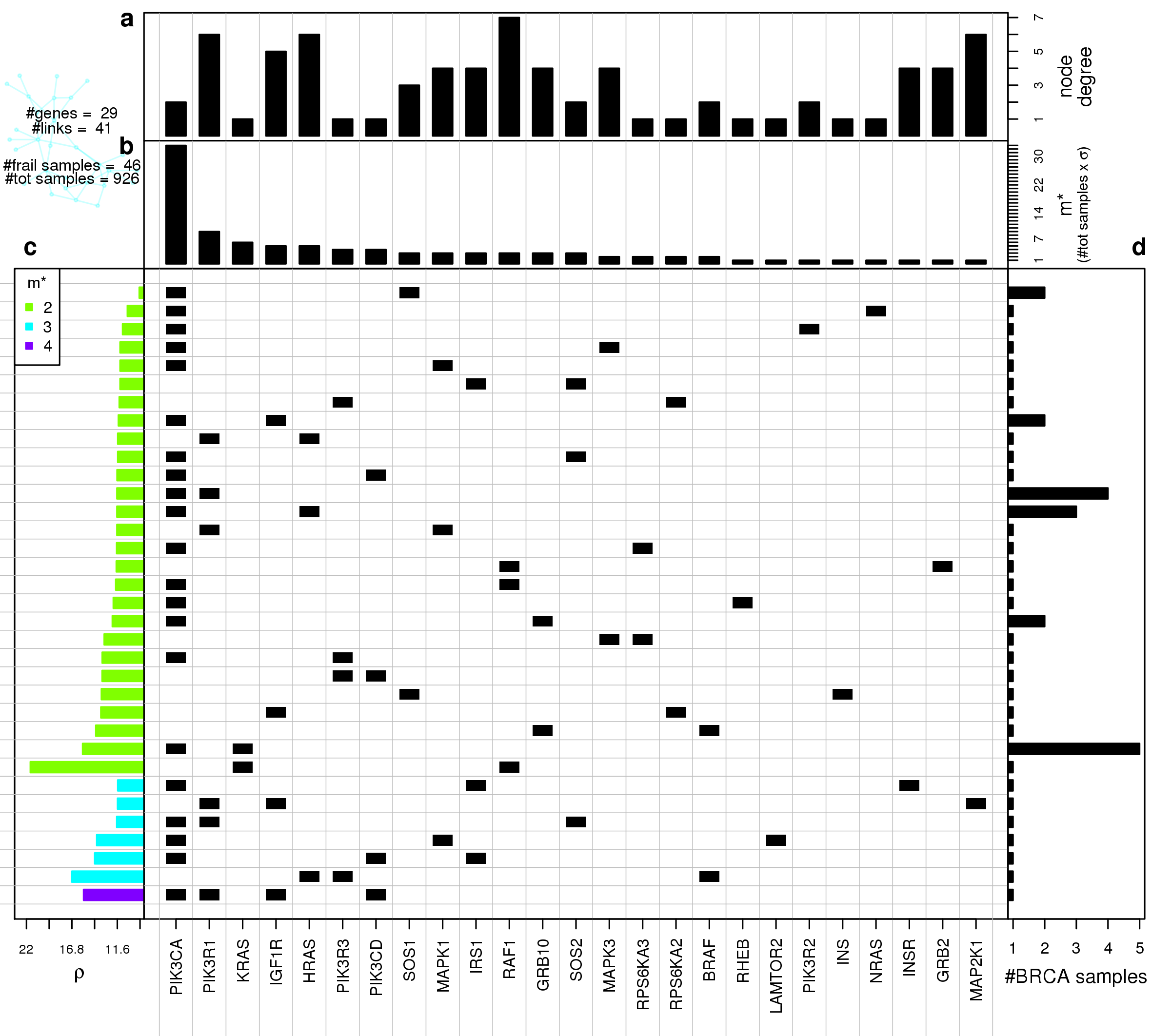
**Supplementary Figure S6**

****

**Figure S6. Components of the GFI for** **MAPK signaling pathway**. On the MAPK signaling pathway gene network we plot the GFI components for each gene of the network that has at least one non-null component. The x-axis is an elementary transformation of the gene co-occurrence σ (*m*∗ = σ · *nsamples* ) and it is displayed in log scale for visualization issues.

**------------------------- mTor signaling pathway -------------------------------**

**Supplementary Figure S7**

****

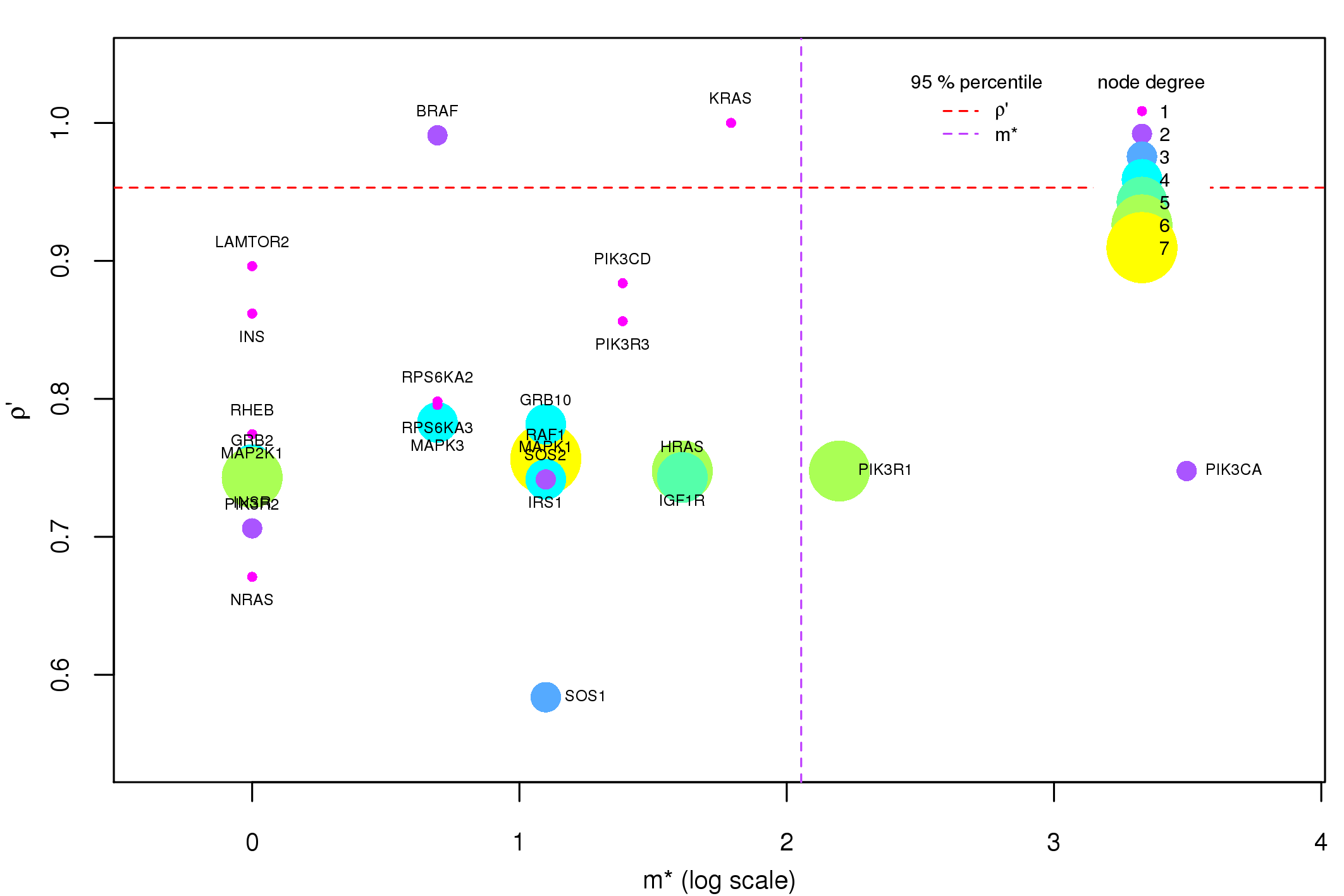
**Figure S7. Gene network frailty analysis for mTor signaling pathway.** Central plot: in each row we plot the somatic mutations co-occurring in the same samples on the mTor signaling pathwaygene network. Rows are ordered as the left-hand side plot, columns are sorted from left to right for decreasing *m*∗. Top panels: from top to bottom is respectively displayed each gene node degree in the target network (a) and the number of patients in which each gene is mutated together with at least another gene (*m*∗) (b); the columns of the central plot are sorted according to decreasing *m*∗. All results were obtained using BC data mapped on KEGG’s mTor signaling pathwaygene network. Left-hand side plot (c): Gene network frailty is plotted for each SM configuration; the ordering on the y-axis is determined in first instance by number of mutations per patient occurring on the gene network, and then in each class the configurations are sorted by ρ. Right-hand side plot (d): for each row of the central plot is displayed the number of patients presenting the corresponding SM configuration.

**Supplementary Figure S8**

****

**Figure S8. Gene-gene co-occurrences and frailty measures for mTor signaling pathway.** For each couple of mTor signaling pathwayco-mutated genes we show the average frailty < ρ > measured on all the BC pathway having the correspondent couple of genes simultaneously mutated using the color scale from green corresponding to null frailty (maximum resilience) to red corresponding to the maximum frailty found in real BC data. The size of each dot is proportional to the log-log scaled number of BC patients presenting the co-mutation.

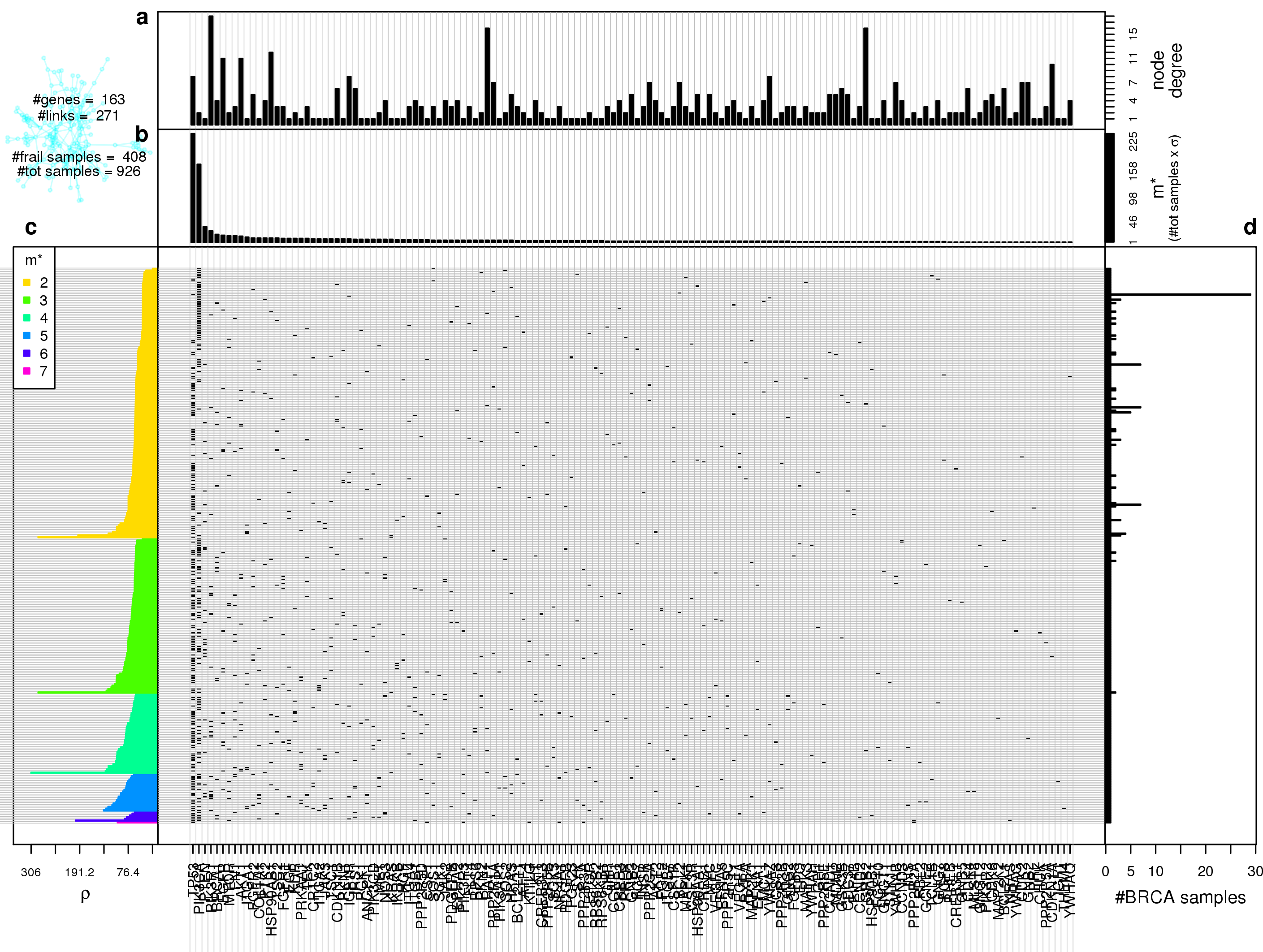
**Supplementary Figure S9**

****

**Figure S9. Components of the GFI for mTor signaling pathway**. On the mTor signaling pathway gene network we plot the GFI components for each gene of the network that has at least one non-null component. The x-axis is an elementary transformation of the gene co-occurrence σ (*m*∗ = σ · *nsamples* ) and it is displayed in log scale for visualization issues.

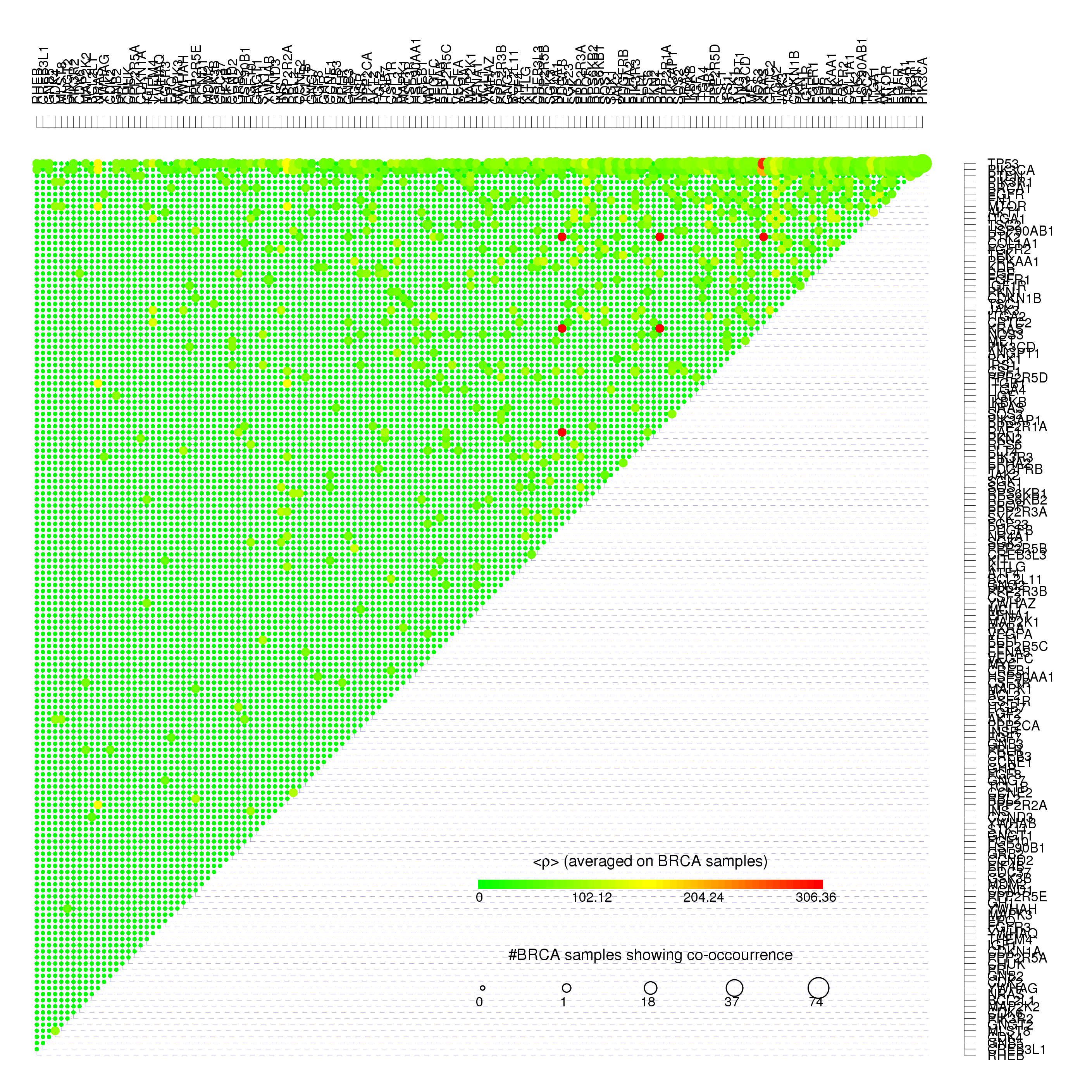
**------------------------- PI3K-Akt signaling pathway --------------------------**

**Supplementary Figure S10**

****

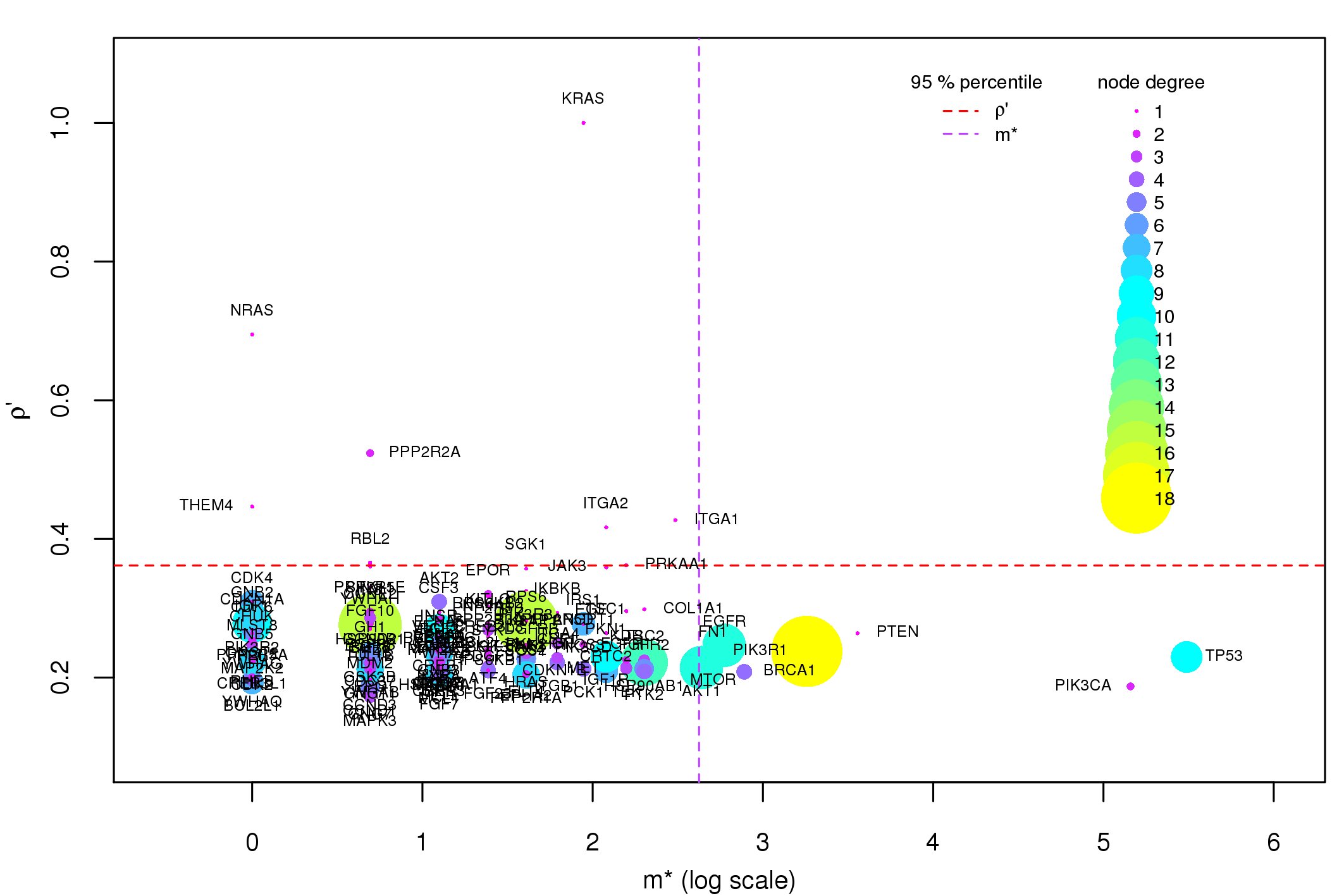
**Figure S10. Gene network frailty analysis for PI3K-Akt signaling pathway.** Central plot: in each row we plot the somatic mutations co-occurring in the same samples on the PI3K-Akt signaling pathwaygene network. Rows are ordered as the left-hand side plot, columns are sorted from left to right for decreasing *m*∗. Top panels: from top to bottom is respectively displayed each gene node degree in the target network (a) and the number of patients in which each gene is mutated together with at least another gene (*m*∗) (b); the columns of the central plot are sorted according to decreasing *m*∗. All results were obtained using BC data mapped on KEGG’s PI3K-Akt signaling pathwaygene network. Left-hand side plot (c): Gene network frailty is plotted for each SM configuration; the ordering on the y-axis is determined in first instance by number of mutations per patient occurring on the gene network, and then in each class the configurations are sorted by ρ. Right-hand side plot (d): for each row of the central plot is displayed the number of patients presenting the corresponding SM configuration.

**Supplementary Figure S11**

****

**Figure S11. Gene-gene co-occurrences and frailty measures for PI3K-Akt signaling pathway.** For each couple of PI3K-Akt signaling pathwayco-mutated genes we show the average frailty < ρ > measured on all the BC pathway having the correspondent couple of genes simultaneously mutated using the color scale from green corresponding to null frailty (maximum resilience) to red corresponding to the maximum frailty found in real BC data. The size of each dot is proportional to the log-log scaled number of BC patients presenting the co-mutation.

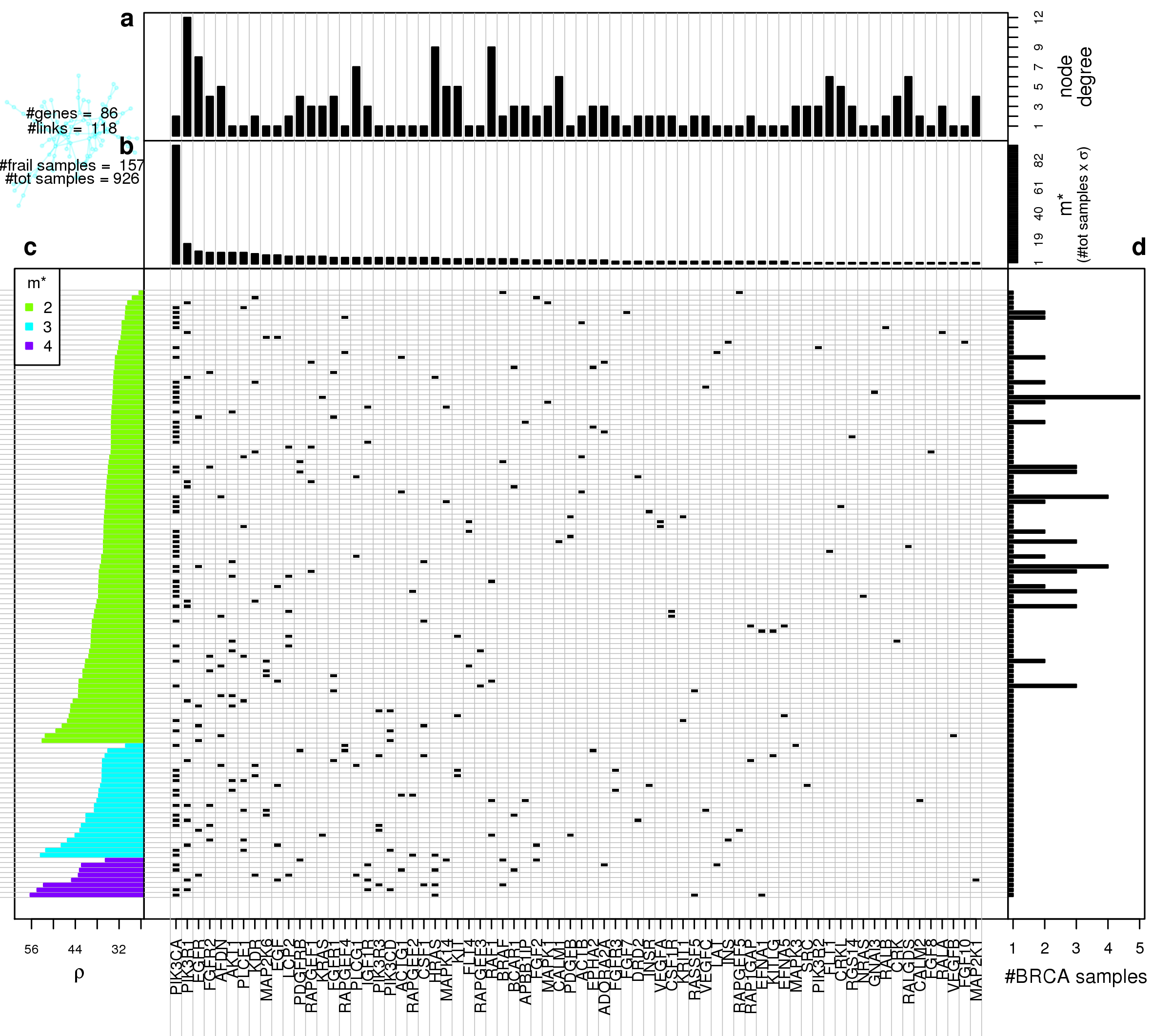
**Supplementary Figure S12**

****

**Figure S12. Components of the GFI for PI3K-Akt signaling pathway**. On the PI3K-Akt signaling pathway gene network we plot the GFI components for each gene of the network that has at least one non-null component. The x-axis is an elementary transformation of the gene co-occurrence σ (*m*∗ = σ · *nsamples* ) and it is displayed in log scale for visualization issues.

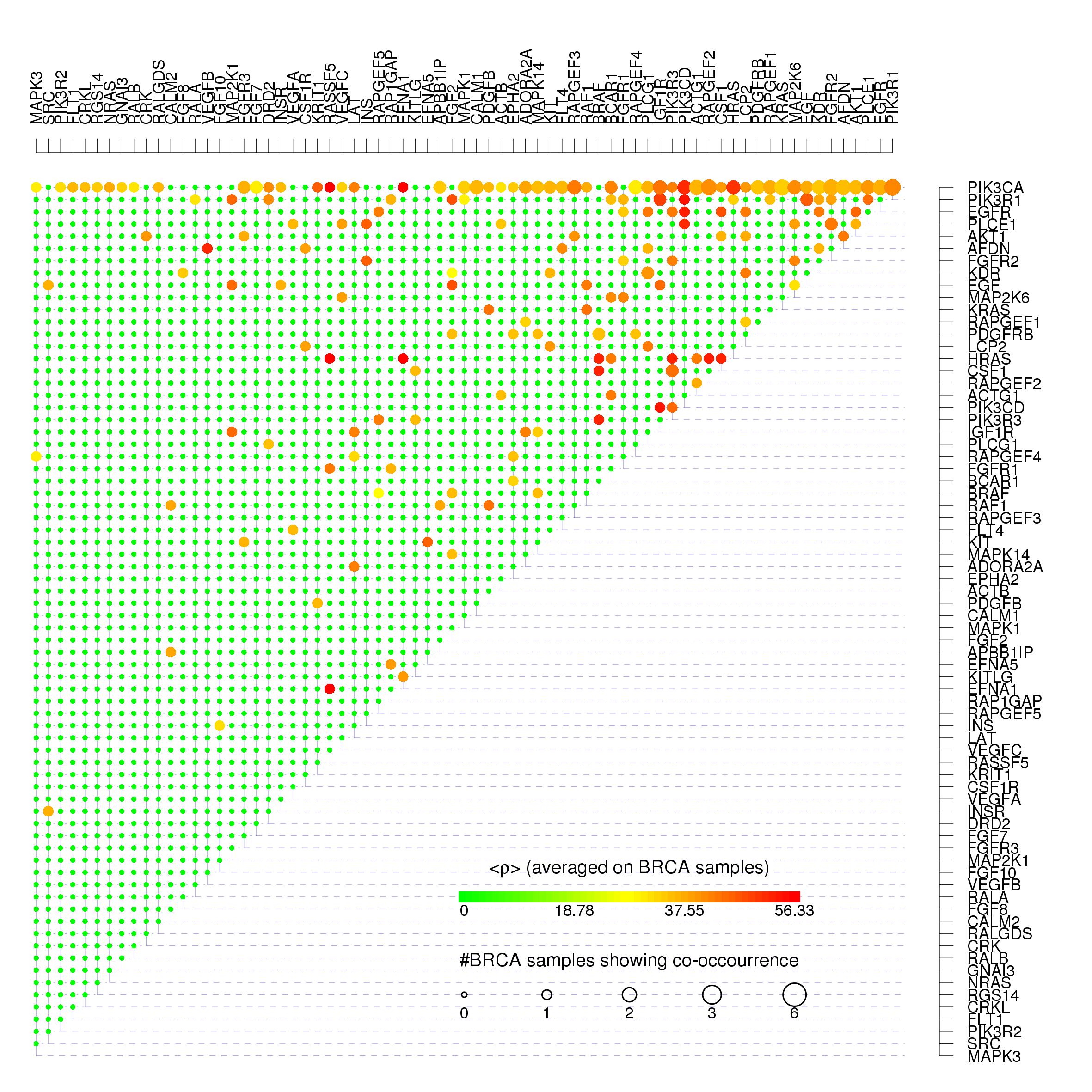
**------------------------- Rap1 signaling pathway --------------------------------**

**Supplementary Figure S13**

****

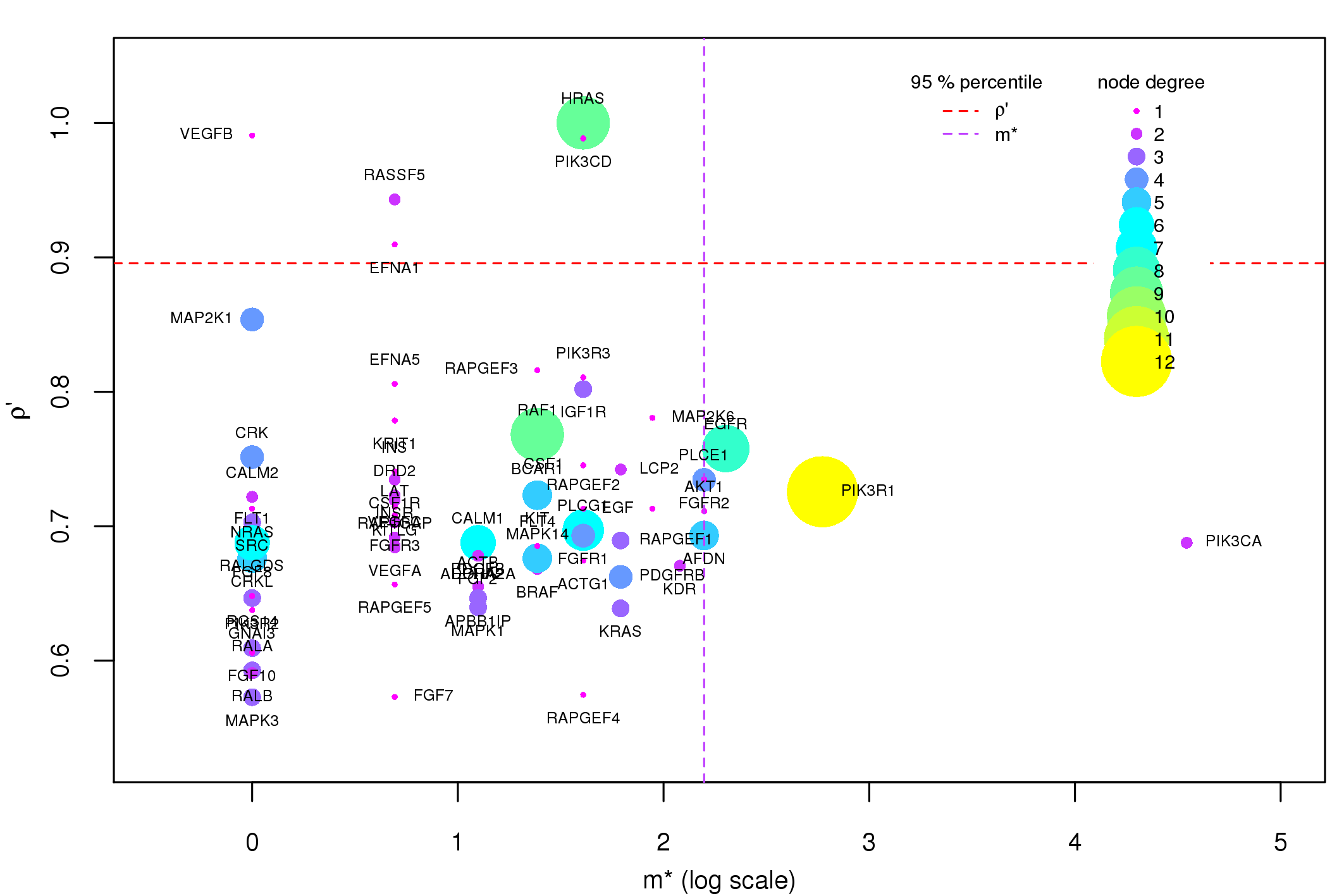
**Figure S13. Gene network frailty analysis for Rap1 signaling pathway.** Central plot: in each row we plot the somatic mutations co-occurring in the same samples on the Rap1 signaling pathwaygene network. Rows are ordered as the left-hand side plot, columns are sorted from left to right for decreasing *m*∗. Top panels: from top to bottom is respectively displayed each gene node degree in the target network (a) and the number of patients in which each gene is mutated together with at least another gene (*m*∗) (b); the columns of the central plot are sorted according to decreasing *m*∗. All results were obtained using BC data mapped on KEGG’s Rap1 signaling pathwaygene network. Left-hand side plot (c): Gene network frailty is plotted for each SM configuration; the ordering on the y-axis is determined in first instance by number of mutations per patient occurring on the gene network, and then in each class the configurations are sorted by ρ. Right-hand side plot (d): for each row of the central plot is displayed the number of patients presenting the corresponding SM configuration.

**Supplementary Figure S14**

****

**Figure S14. Gene-gene co-occurrences and frailty measures for Rap1 signaling pathway.** For each couple of Rap1 signaling pathwayco-mutated genes we show the average frailty < ρ > measured on all the BC pathway having the correspondent couple of genes simultaneously mutated using the color scale from green corresponding to null frailty (maximum resilience) to red corresponding to the maximum frailty found in real BC data. The size of each dot is proportional to the log-log scaled number of BC patients presenting the co-mutation.

**Supplementary Figure S15**

****

**Figure S15. Components of the GFI for Rap1 signaling pathway**. On the Rap1 signaling pathway gene network we plot the GFI components for each gene of the network that has at least one non-null component. The x-axis is an elementary transformation of the gene co-occurrence σ (*m*∗ = σ · *nsamples* ) and it is displayed in log scale for visualization issues.