Supplementary material

Hunt for the tipping point during endocrine resistance process in breast cancer by dynamic network biomarkers

Rui Liu, Jinzeng Wang, Masao Ukai, Ki Sewon, Yutaka Suzuki, Pei Chen, Haiyun Wang, Kazuyuki Aihara, Mariko Okada-Hatakeyama, and Luonan Chen

Supplementary method

Landscape dynamic network biomarker and data processing

Given a biomolecular interaction network structure, an efficient method to detect DNB, called the landscape dynamic network biomarker, is proposed by employing the local-landscape method on the basis of the three DNB statistic properties. Specifically, first we mapped the genes to the gene regulation network (e.g., an integrated STRING network for *Homo sapiens*). Second, the network was partitioned into many local networks. Each local network contained a center node/gene and all of its first-order neighbors based on the network structure. The index I_{local}-score at time point t for a local network with N members (i.e., one center node and N-1 first-order neighboring nodes) was then calculated through the following definition:

$$I_t = |\Delta SD_t(in)| \, [\, |\Delta PCC_t(in)| + |\Delta PCC_t(out)| \,],$$

where

$$|\Delta SD_t(in)| = \frac{\sum_{i=1}^{N} |SD_t(gene\ i) - SD_{t-1}(gene\ i)|}{N}$$

is the average differential standard deviation (in absolute value) of the nodes inside the local network;

$$|\Delta PCC_t(in)| = \frac{\sum_{i=1,j=1}^{N} |PCC_t(gene\ i,\ gene\ j) - PCC_{t-1}(gene\ i,\ gene\ j)|}{N*N}$$

is the average differential Pearson's correlation coefficient (in absolute value) inside the local network, i.e., both gene i and gene j are in the local network;

$$|\Delta PCC_t(out)| = \frac{\sum_{i=1,j=1}^{N} |PCC_t(gene\ i,\ gene\ j) - PCC_{t-1}(gene\ i,\ gene\ j)|}{N*N}$$

is the average differential Pearson's correlation coefficient (in absolute value) between a member (gene i) in the local network and those (gene j) outside.

Theoretically, when the system approaches the tipping point, i.e., $t \in critical\ state, and\ t-1 \notin critical\ state$, there are three cases for the local networks:

(1) In the local network, all the members (or nodes) are DNB genes;
(2) In the local network, there are DNB and non-DNB genes;
(3) In the local network, all the members are non-DNB genes.
According to the three cases respectively, there are critical behaviors shown as in the following table:

Case	Members/nodes		SD_t	$|\Delta SD_t(in)|$	$PCC_t(in)$	$|\Delta PCC_t(in)|$	$PCC_t(out)$	$|\Delta PCC_{out}(t)|$	I_t
(1)	All DNB		↗	↗	↗	↗	D ↗	↗	↗
							N ↘		
(2)	DNB & non-DNB	D	↗	↗	D ↗	↗	D ↗	↗	↗
					N ↘	↗	N ↘	↗	
		N	→	0	D ↘	↗	D ↘	↗	
					N →	0	N →	0	
(3)	All non-DNB		→	0	→	0	D ↘	↗	0
							N →	0	

Notation: the system is near a tipping point, i.e., it moves from time point *t*-1 to *t*, with $t \in$ *critical state, and t-1 $\notin$ critical state*.
1. "↗" represents the increase of the index; "↘" represents the decrease of the index; "→" represents that there is no significant change in the index;
2. "**D**" stands for the DNB genes, or the PCC with DNB genes; "**N**" stands for the non-DNB genes, or the PCC with non-DNB genes
3. SD_t is the standard deviation at time t; $PCC_t(in)$ is the Pearson's correlation coefficient between two genes inside the local network; $PCC_t(out)$ is the Pearson's correlation coefficient between genes inside the local network and genes outside.
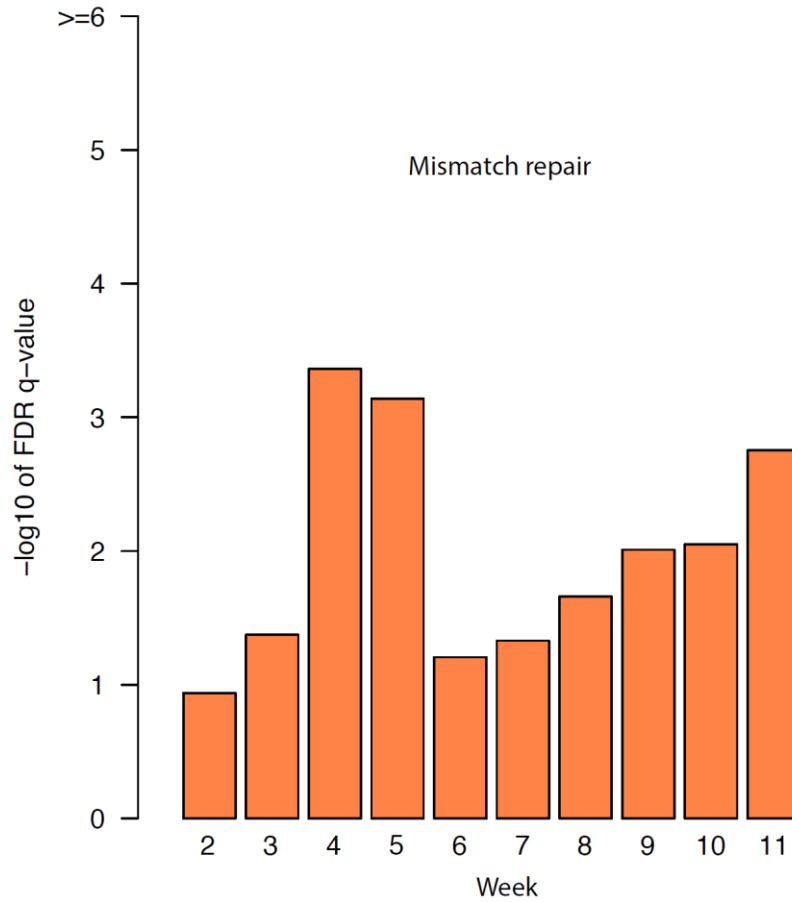
Thus, the index I_{local}-score, I_t, can quantitatively characterize the criticality of the state for each node or gene. Clearly, each gene/node has an I_t value, and hence those I_ts for all of genes with the time evolution construct a landscape as shown in Figure 4B. When the system approaches the critical state, I_t increases drastically based on the three statistic conditions of DNB. To quantify the collective behavior, we chose the top-m genes based on their I_t as the DNB members, and the average value I_t of the top-m genes/nodes as the criterion to quantify the tipping point or the critical state.
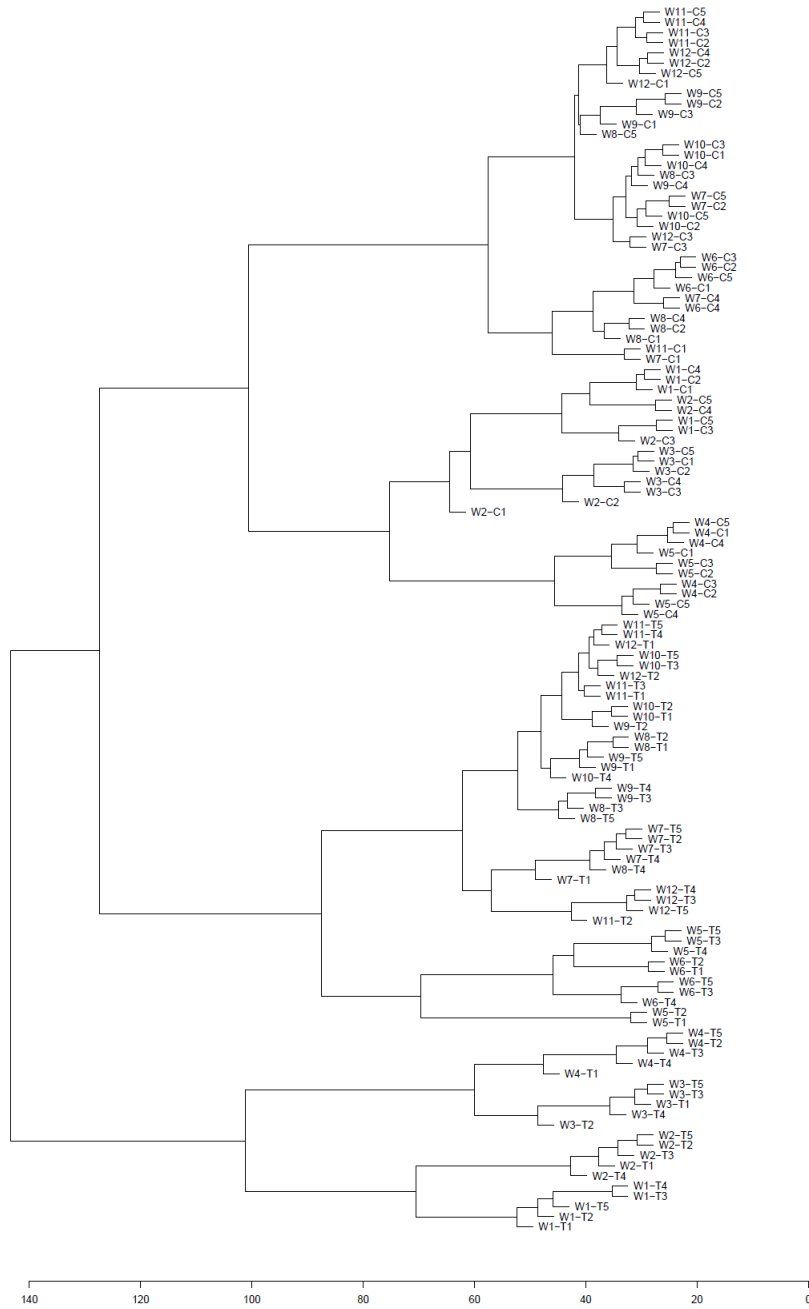
Supplementary Figure S1. The detailed cluster of 120 samples for RNA-seq data.

T represents tamoxifen treated (TamR) sample while C indicates control sample. W indicates week. The number from 1 to 5 following after T/C stands for different replicate.

Supplementary Figure S2. Cell cycle pathway is ectopically activated in the resistance cells.

Proteins encoded by the upregulated genes in week 5-12 compared with week 1-4 are shown in red.

Supplementary Figure S3. The mismatch repair pathway is statistically associated with the 4th/5th week, with the great value of –log10(FDR q-value).
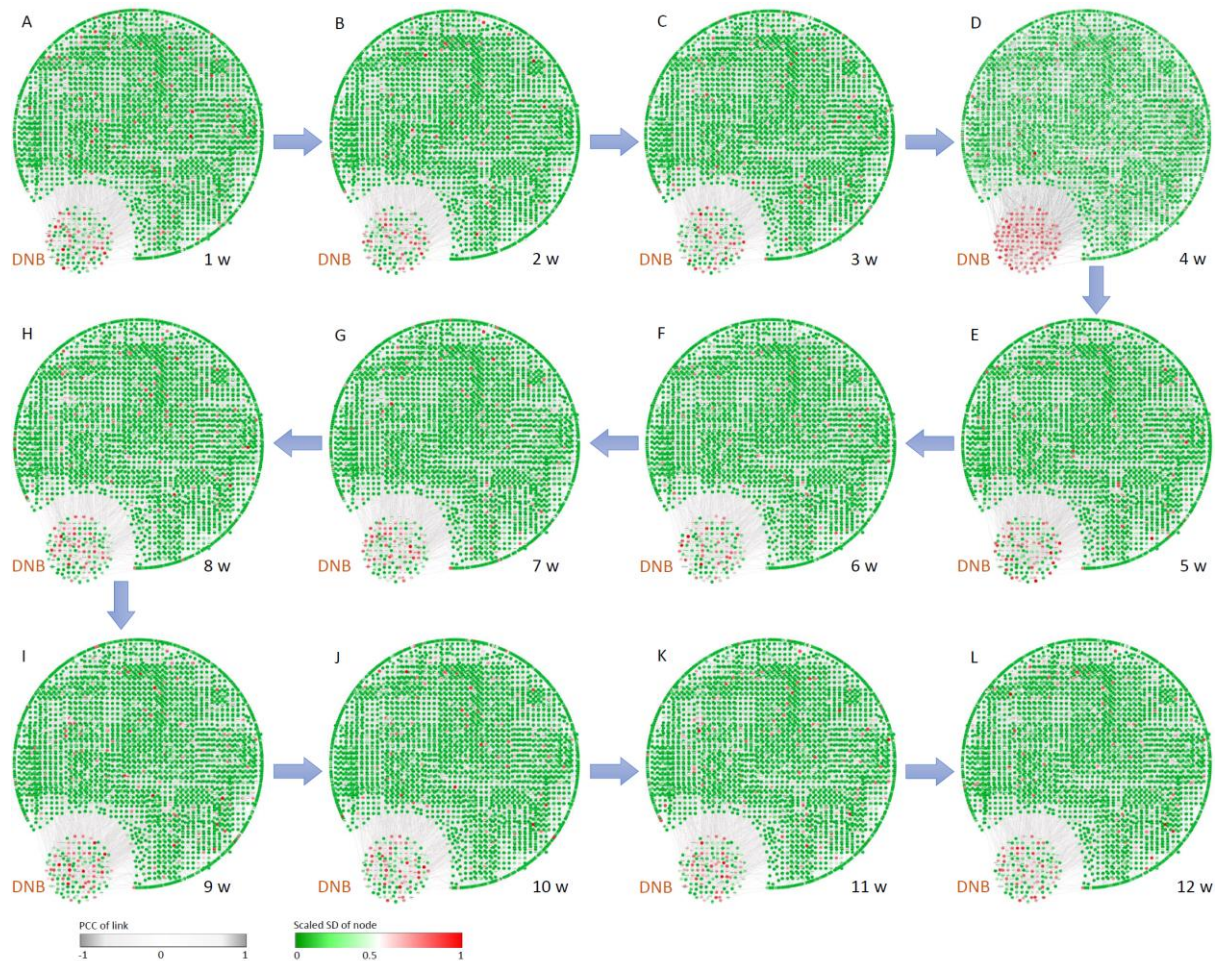
Supplementary Figure S4. Dynamical changes in the overall human PPI network.
Dynamical changes in the overall human PPI network including the DNB (227 genes located at the left bottom corner) we identified. For MCF-7 human breast cancer cells, we determined the dynamical evolution of the network structure of the entire Homo molecular interaction network (protein-protein interactions and TF-target regulations), including the DNB module/subnetwork, in terms of the landscape of DNB. (A)-(D) Molecular networks during the non-resistance stage. (E) Molecular network during the pre-resistance stage. (F)-(L) Molecular networks during Tamoxifen resistance stage. Clearly, the sampling time point 4 weeks after Tamoxifen exposure (4 w) was the tipping point according to our criterion, which corresponded to the pre-disease or pre-deterioration state when the genes of the DNB network behaved significantly differently from others.
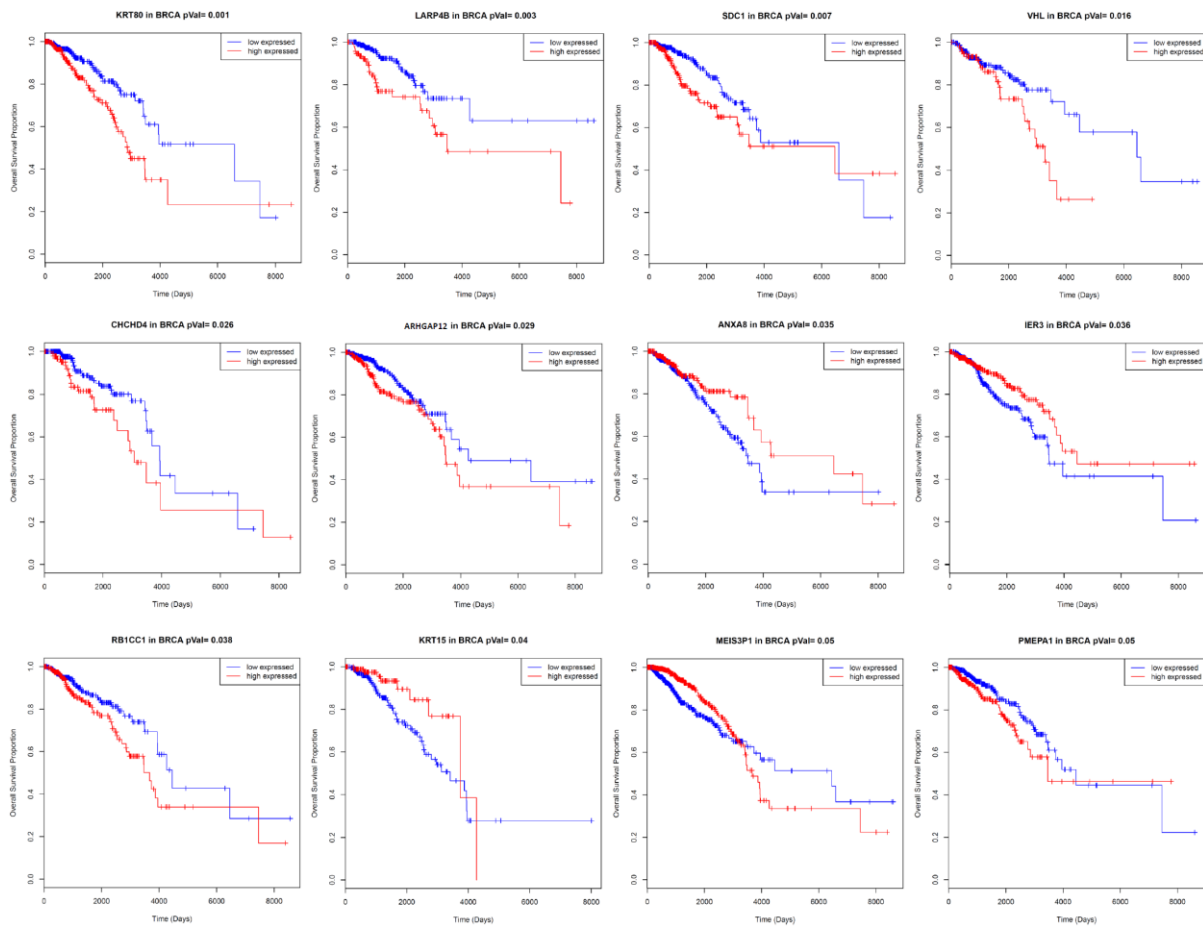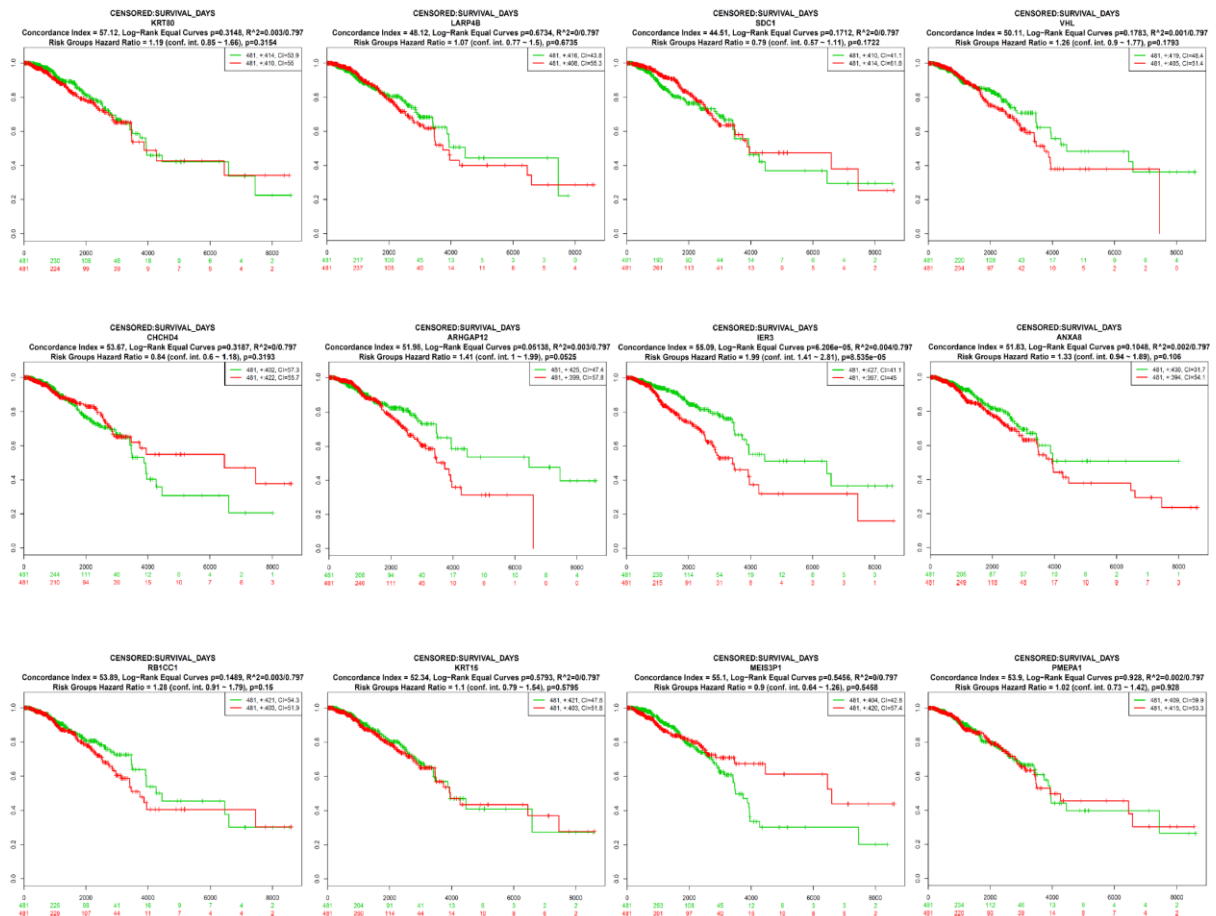
Supplementary Figure S5. Survival analysis based on single genes of DNB in EBI dataset.

We listed the survival analysis results with p-values<0.05 on the basis of some single genes of DNB. We divided the samples into two groups based on the median of gene expression level. Each group respectively included 496 samples. It can be seen that most survival analyses are almost indistinguishable, while the DNB genes together (see Fig. 7 in the main text), through a Cox regression model, and are capable to classify the samples into two groups with distinct prognosis.

The survival analysis based on 962 samples of TCGA data

Supplementary Figure S6. Survival analysis based on single genes of DNB in TCGA dataset.

We listed the survival analysis results with p-values<0.05 on the basis of some single genes of DNB. We divided the samples into two groups based on the median of gene expression level. Each group respectively included 481 samples. It can be seen that most survival analyses are almost indistinguishable, while the DNB genes together (see Fig. 7 in the main text), through a Cox regression model, and are capable to classify the samples into two groups with distinct prognosis.
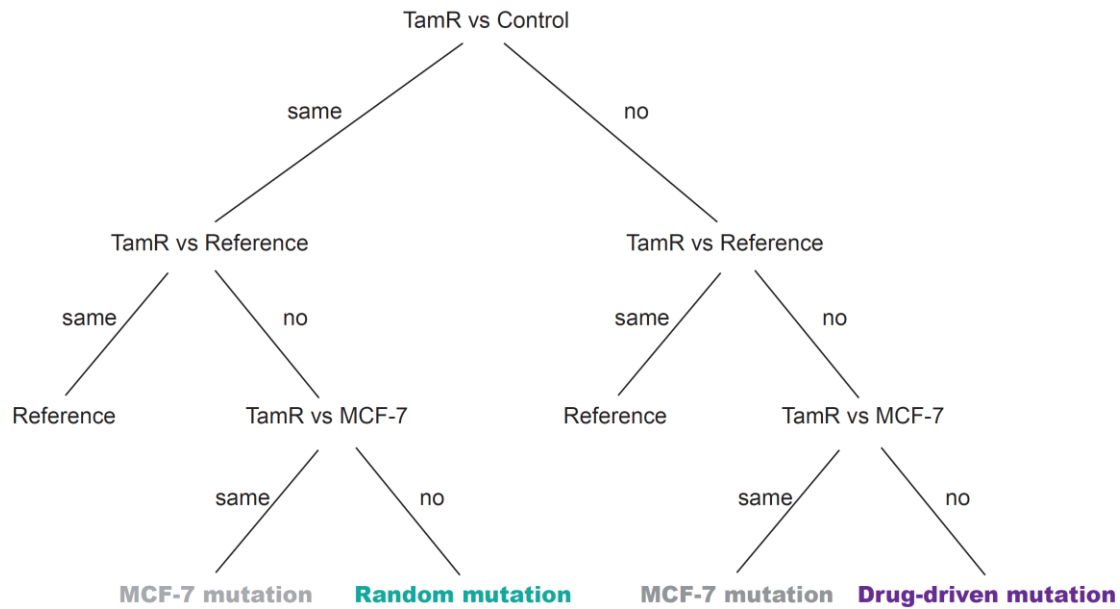
Supplementary Figure S7. Mutation classification.

For each mutation site detected by GATK, we compared its genotype in the TamR samples with the genotype respectively in control samples, in reference genome and in MCF-7 cell line, to determine the mutation type. By the multiple comparisons of the genotypes above, the mutations were further classified as drug-driven mutation, random mutation, and MCF-7 mutation
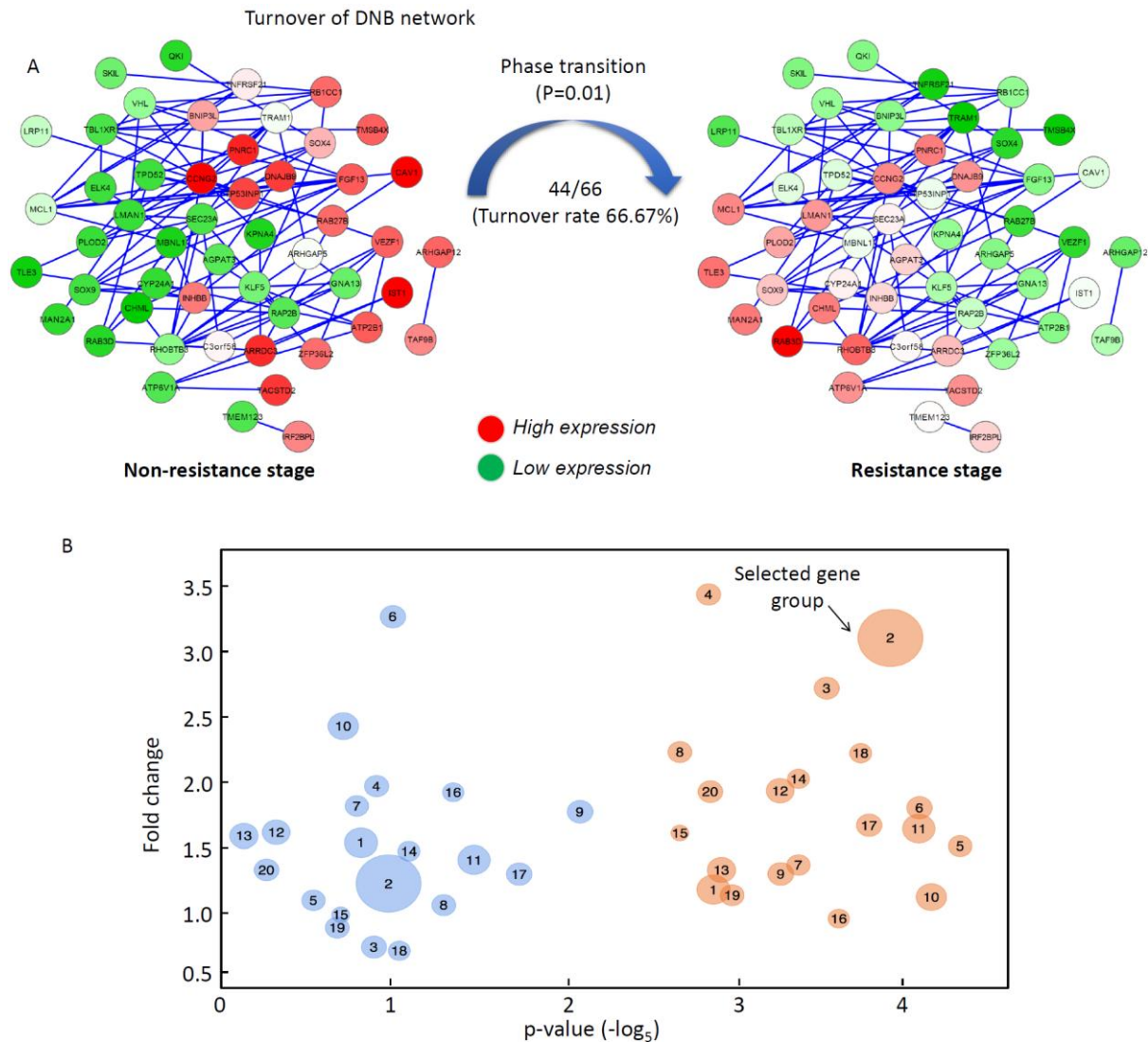
Supplementary Figure S8. The turnover of the DNB network.

(A) The DNB network is generated by mapping the DNB genes and their 1st-order neighbors onto the STRING network. In this network, over 66.67% genes have reversal (or turnover) expressions when the system progresses from the non-resistance state to the resistance state by passing the pre-resistance state (the tipping point), comparing with only 19.41% turnover ratio for all genes. (B) The bubble chart shows the statistical significance of the identified DNB, where each bubble presents a cluster group with an order number, and the size of the bubble corresponds to the number of genes in the group. It is seen that the identified DNB genes (the 2nd clustering group in orange) are statistically significant with p-value 2.2E-03 and over 3-fold change comparing with the control group at the 4th week, which is more significant than those at the 2nd week (left groups in blue).

Supplementary Table S1. The top 1000 differentially expressed genes in week 5-12 compared with week 1-4.

Supplementary Table S2. The list of DNB genes and the upstream transcriptional factors.

Supplementary Table S3. Turnover genes of the DNB network (including DNB genes and their first-order neighbors from STRING network).

Supplementary Table S4. The mutation sites exist at least in the 4 time points across 12 weeks, including those occurring from the 4th/5th week and existing at least in the 4 time points thereafter.

Supplementary Table S5. 15 genes with drug-driven mutations occurring in the exons.