Supplementary Figures

Dysfunction of PLA2G6 and CYP2C44 associated network signals imminent carcinogenesis from chronic inflammation to hepatocellular carcinoma

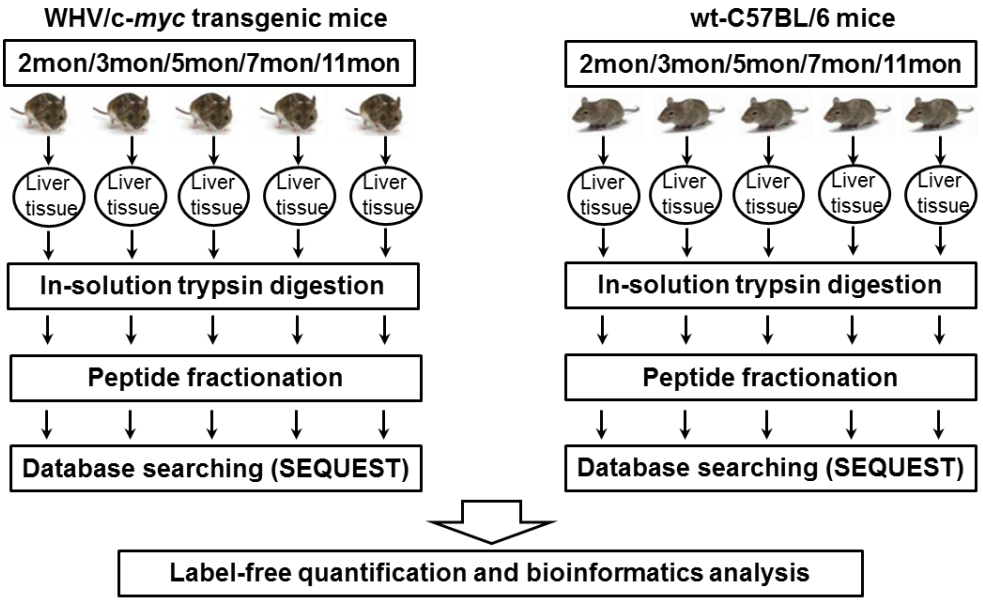

Figure S1
Workflow of label-free proteomics.

A schematic chart shows the detailed steps of proteomics by label-free quantified strategy combined with offline peptide fractionation and liquid chromatography–tandem mass spectrometry (LC-MS/MS) for liver samples of 50 mice. Meanwhile, this transgenic mouse carrying the c-*myc* oncogene can be successfully induced to develop HCC under control of woodchuck hepatitis virus (WHV) DNA sequences. Our previous study has reported some insights and indicators for early hepatocarcinogenesis based on proteomic mining in the dysplastic liver of WHV/c-myc mice [1].

All collected liver tissue were individually cut into small pieces at first and washed three times with cold RPMI buffer (Sigma). Then the tissue sample was homogenized in cold lysis buffer (8 M urea, 4% (v/v) CHAPS, 40 mM Tris base) and sonicated for 3 minutes. Afterwards, liver tissue sample was centrifuged under 4℃, 15,000 g for 60 minutes to remove insoluble fraction and transferred to a new tube. Finally, the protein concentration of liver tissue sample was quantified using Bradford kit (Bio-rad).

For label-free tests, the liver lysates of 200 μg from 5 mouse individuals at each stage of each type mouse were applied for protein digestion. For TMT labeling, the liver lysates from 3 mouse individuals (With similar SDS-PAGE commassie blue staining pattern) at each stage of each type mouse were mixed equally to a total amount of 200 μg. Filter aided sample preparation (FASP) was used for proteolysis [2]. Briefly, samples were loaded on the 10kDa filter units (1.5μl tube), and washed twice with 200μl urea buffer (8M urea, 100mM Tris, pH8.5) at 14000g for 30min, to remove SDS. Then added 100μl of 50 mM iodoacetamide (IAA) diluted in urea buffer, and incubated for 30 min in the dark. The alkylated proteins were washed with 100μl urea buffer and 50mM NH_4HCO_3, each for three times. Finally, added 40μl 50 mM NH_4HCO_3 buffer with trypsin (the enzyme-to-protein ratio is 1:50) into the sample and incubated at 37℃ for 18h. For TMT labeling experiment, protein lysates were changed into and digested with trypsin in 200 mM tetraethylammonium borohydride (TEAB). The peptides were harvested by centrifugation, then lyophilized.

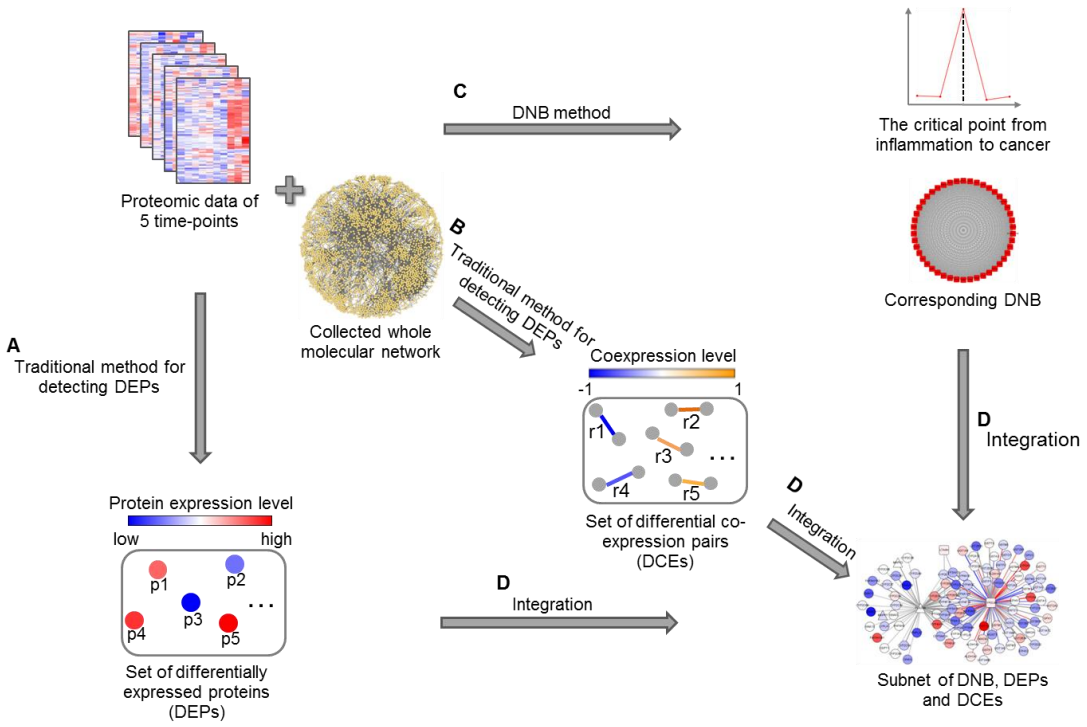

Figure S2
Flowchart of computational analyses.

A. Total 1465 DEPs (table S2) were identified by the following method.

To identify disease-associated dysfunctions of proteins, regulations and biological processes along hepatocarcinogenesis, we performed differentially expressed protein (DEP) analysis as well as differential coexpression (DCE) analysis (Figure S3, A and B). These disease-associated differences were compared not only between cases and controls at each same stage but also among cases at two different stages. Differentially expressed proteins were identified by Welch's t-test with two-tailed p-value less than 0.05.

$$t = \frac{z}{\sqrt{V/v}}$$

where n is the same total number of transgenic mice or control mice at each time point; $\overline{X_1}$ and s_1^2 are mean and variance of transgenic mice at time point t (t = 1, … , 5), respectively; $\overline{X_2}$ and s_2^2 are mean and variance of control mice at t or transgenic mice at t+1 (t = 1, … , 4).

B. Total 3338 DCEs (table S3) were identified by the following method.

Similar to DEP analysis, differential coexpression protein-pairs were measured by the difference of Pearson correlation coefficients (ΔPCC) between case and control at each same stage as well as among cases at two different periods. Firstly, according to the knowledge-based background network, we only considered total 6631 binary interactions whose nodes all have effective quantitative values. For this background network of mouse, we collected 48374 binary interactions without redundance from public databases (e.g. KEGG [3], BioGRID [4] and TRED [5]). Types of these interactions include protein-protein interactions (e.g. direct interactions into protein complexes and indirect signaling transmissions), protein-DNA interactions (e.g. transcription factors-involved regulatory interactions). Here, we just considered total 6631 binary interactions, whose nodes both belong to the 3,372 proteins with effective quantitative values. Next, in order to determine the corresponding background distributions of ΔPCCs for each pairwise period, we firstly assumed ΔPCCs population

pool of random protein-pairs among 3372 proteins at each different comparative conditions to estimate the corresponding population parameters (i.e. mean, standard deviation and degree of freedom for the population). Then, corresponding ΔPCCs of 6631 known protein-pairs were z-transformed according to the previous parameters for each corresponding condition. Further, we filtered significant DCEs based on the t-distribution with two-tailed p-value less than 0.05.

$$t = \frac{z}{\sqrt{V/v}}$$

where z is the normalized ΔPCC of each link; V is estimated by $(n-1)\frac{s^2}{\sigma^2}$, where n is the total number of protein-pairs,s^2 is the unbiased estimate of variance from 6631 links, and σ^2 is the expected variance.

C. Total 48 DNB members and its associated TFs were identified by DNB analysis (The details of its algorithm was described in MATERIALS AND METHODS and Figure S4).

D. Due to that complex diseases result from abnormalities in the systematic interplay of multiple molecules even biological processes [6-8], integrating relatively complete interactome of corresponding organism is necessary to comprehensively understand pathological mechanisms of hepatocellular carcinoma at molecular level.

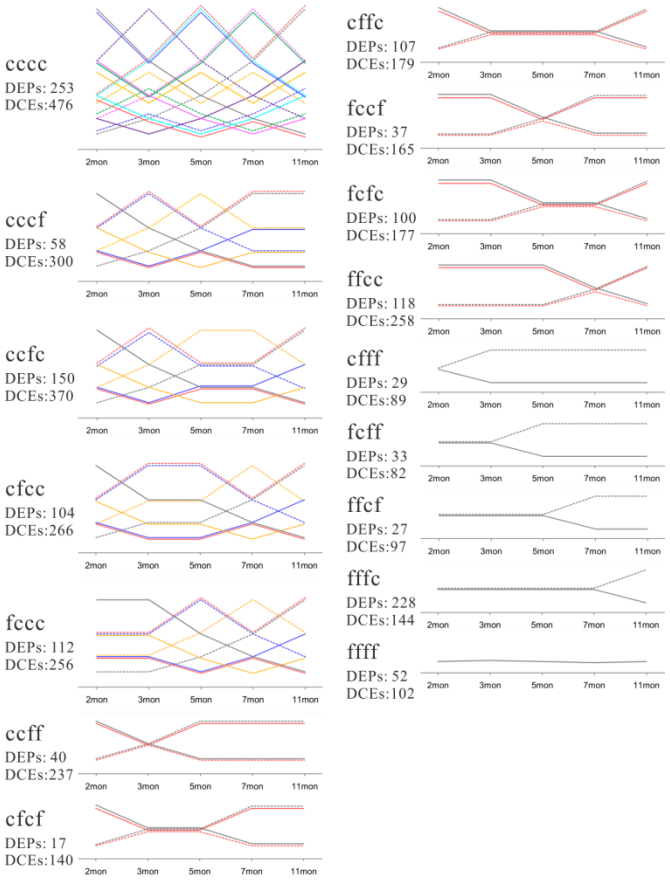

Figure S3
16 different trend patterns of time course analysis.

A series of curve diagrams visually represent dynamic trends of DEPs and DCEs. To characterize different states in the initiation and progression of HCC by trend changes of traditional biomarkers, we classified the DEPs and DCEs into 16 different patterns based on their trends for two consecutive time points for transgenic mice. For each differentially expressed protein, if the maximum ratio of their expression level for transgenic mice between two consecutive time points is more than 1.3, we consider this protein to have a change at the corresponding period. Also, for each differential coexpression protein-pair, the threshold of change is more than 0.4 at its absolute difference of Pearson correlation coefficient (ΔPCC) for the transgenic mice between two consecutive time points. Here, we used “c” and “f” to represent separately the changed and the unchanged at expression (or coexpression) level in two consecutive time points. For instance, ccfc for DEPs means that there are expression changes (c) in 2-3, 3-5, 7-11 months and there are no expression changes (f) in 5-7 month.

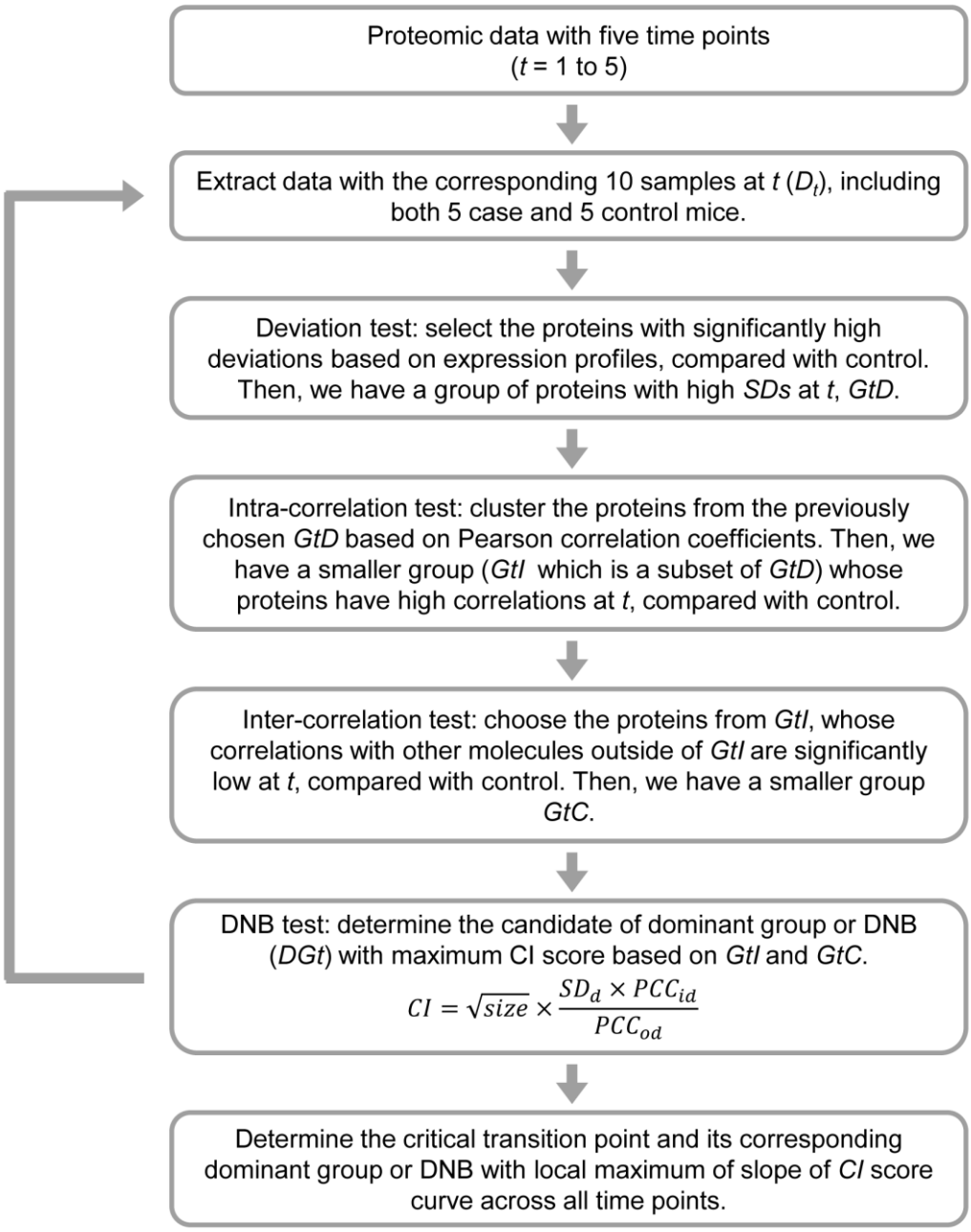

Figure S4
Algorithm of DNB-detected method.

A flowchart describes the detailed steps to determine DNB based on our mathematical method.

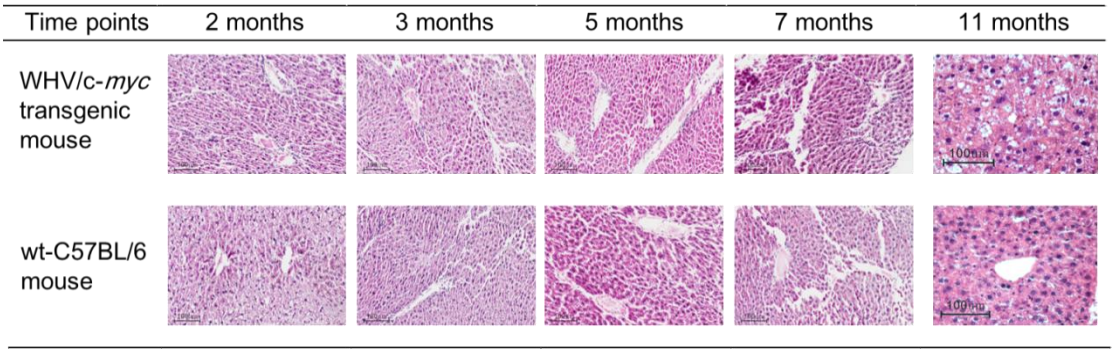

Figure S5
Hematoxylin and eosin (H&E)-stained histological sections of livers.

Histological images of H&E-stained liver sections show different morphological details between WHV/c-*myc* transgenic mice and corresponding wild-type mice at five different time points, respectively. Here, mice livers were excised and fixed immediately in Immunol Staining Fix Solution (Beyotime, P0098) at room temperature for 3.5 h. The tissues were then embedded in paraffin, cut into sections (5 μm thick). Following deparaffinization and rehydration, the slices were stained with a Hematoxylin and Eosin Staining Kit (Beyotime, C0105) to assess liver histology and morphology. Except two images of 11 months at 40× objective magnification and 10× ocular magnification, the others were captured at 10× magnification on Aperio Scan Scope AT Turbo instrument (Leica Biosystems, Buffalo Grove, IL).

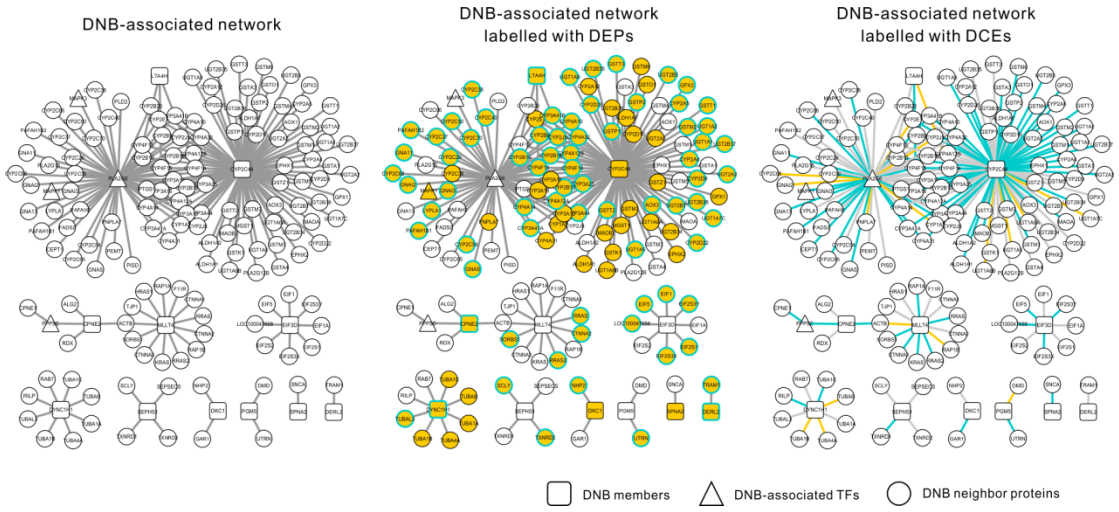

Figure S6
DNB-associated network.

A. This DNB-associated network was constructed by integrating DNB members and their first-order neighbors, according to the knowledge-based molecular network of mouse.
B. DEPs were filled with orange on corresponding nodes of the network. Especially, colored nodes with cyan border represent DEPs with changes during 3 to 7 months of age.
C. DCEs with changes during 3 to 7 months of age were colored cyan. Else, the other DCEs were colored orange.

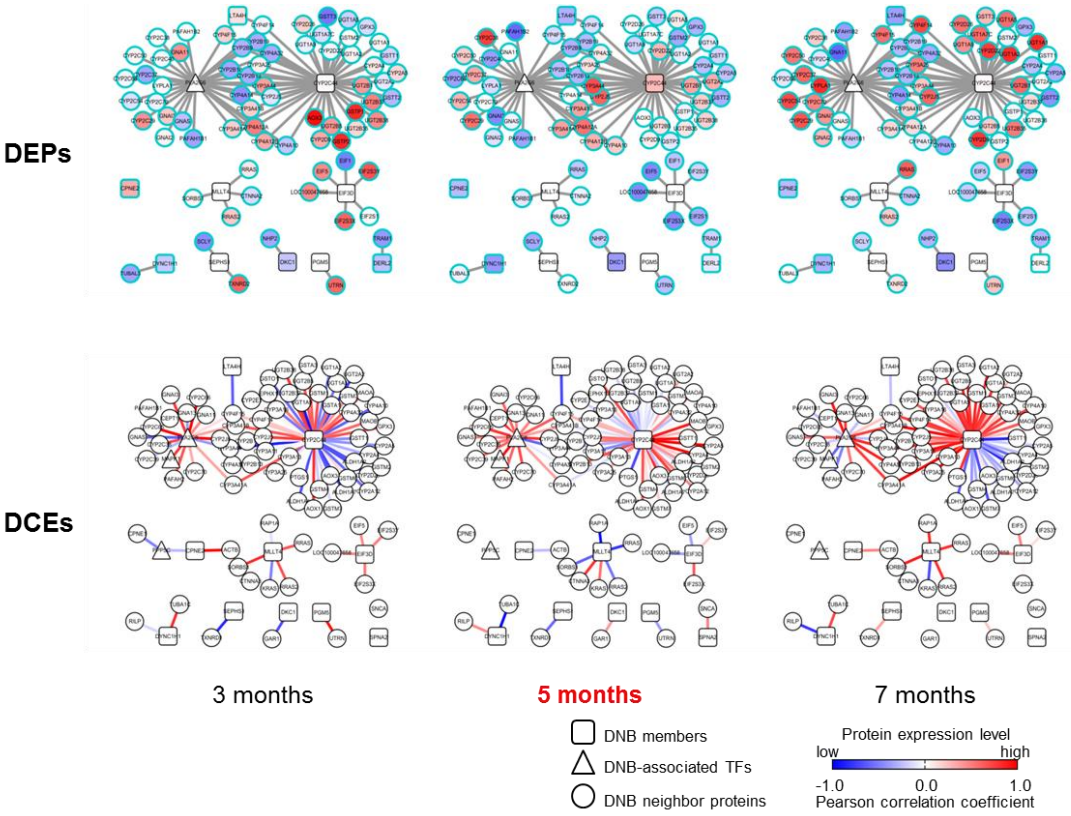

Figure S7
Rewiring of DNB-associated network and dynamic changes of DEPs and DCEs during critical period for wt-C57BL/6 mice.

This series of networks separately demonstrate the dynamic changes or rewiring of DNB-associated network in expressions (DEPs) and coregulations (DCEs) during the critical period.

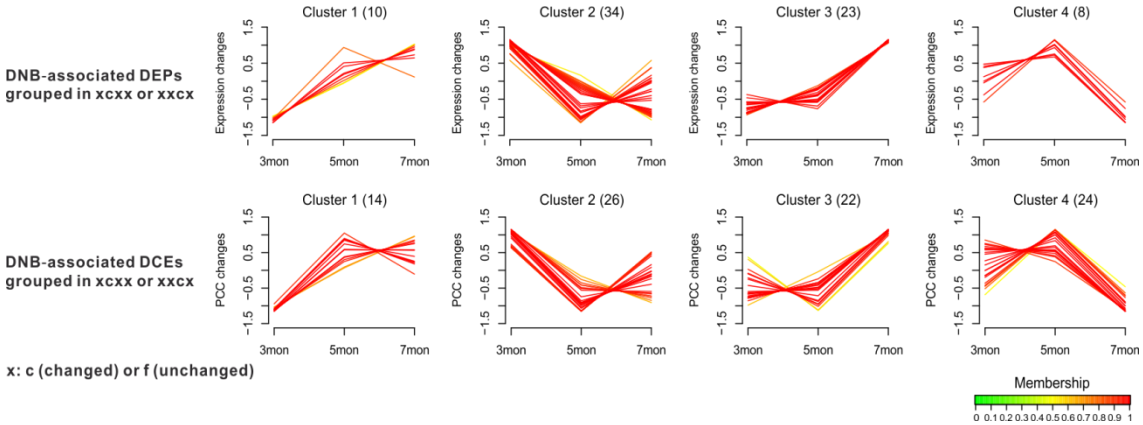

Figure S8
Classifying DNB-associated DEPs and DCEs whose expressions or coexpressions change from 3 to 7 months by Mfuzz.

The dynamical changes of DNB-associated DEPs and DCEs were classified by Mfuzz [9] in R into variant change (dynamic) patterns during two consecutive periods (e.g. 3 to 5 months and 5 to 7 months), which can clearly demonstrate severity-dependent dysfunctions from inflammation to cancer initiation. Obviously, clusters 1 and 2 have respectively up-regulated and down-regulated changes before the critical point (5-month-old for transgenic mice), while Clusters 3 and 4 have respectively up-regulated and down-regulated changes after the critical point.

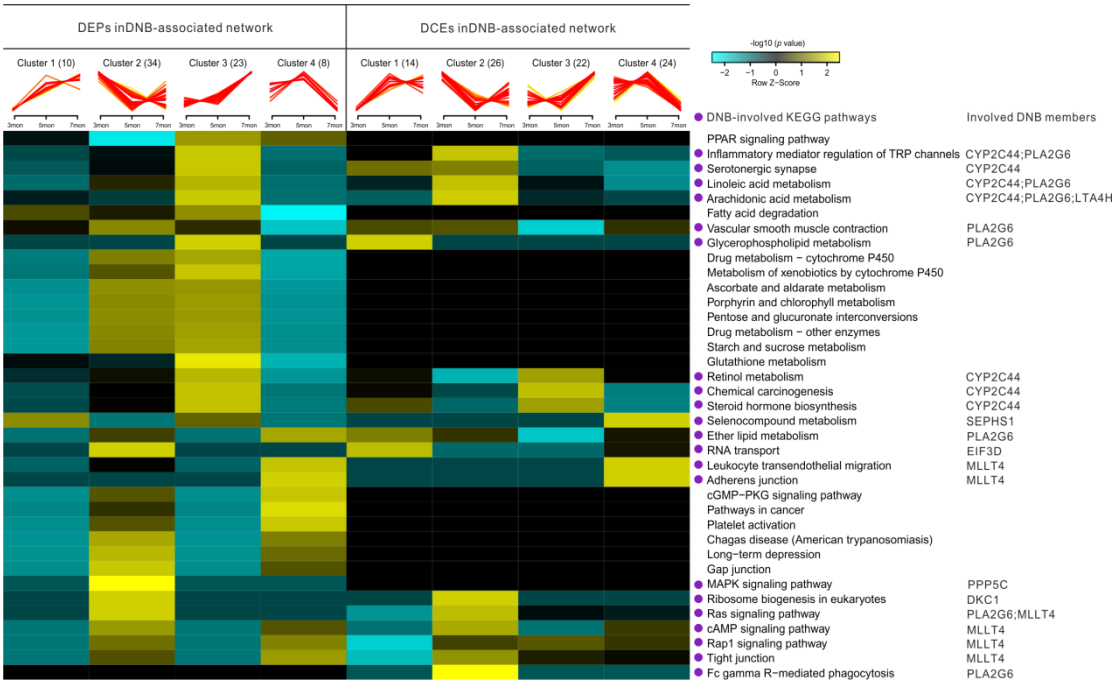

Figure S9
Functional analysis of dynamic patterns by DNB-associated DEPs and DCEs.

Functional analyses on DNB-associated DEPs or DCEs are conducted separately. After comparing the enriched pathways of DEPs with ones of DCEs, we found most of functional phenotypes of DNB-associated DEPs are similar to the overall descriptions but DCEs are more specifically annotated by only 17 biological functions (including lipid metabolisms, cell communication, signal transduction, immunity-associated pathways, etc.) and DEPs are the same to members of whole network except ras signaling pathway and Fc gamma R-mediated phagocytosis. Interestingly, we found that some dysfunctions enriched by both of DEPs and DCEs (e.g. arachidonic acid, linoleic acid and ether lipid metabolisms and inflammatory mediator regulation of TRP channels) appear earlier in DCEs before the critical period than in DEPs after the critical period, which indicates that the disease-related dysregulations are strongly associated with abnormities in molecule expressions.

In addition, from dysregulations of antagonism or coordination for some of these above pathways (e.g. arachidonic acid and linoleic acid metabolisms and inflammatory mediator regulation of TRP channels), DCEs have the down-regulated pattern but DEPs have the up-regulated pattern; for some other pathways (e.g. ribosome biogenesis in eukaryotes, RNA transport, cAMP signaling, cell communication), DCEs and DEPs have the same regulated patterns. Meanwhile, for some pathways enriched by DEPs (e.g. carbohydrate and lipid metabolisms, cell communication and etc) and three pathways enriched by DCEs (e.g. linoleic acid and ether lipid metabolisms, steroid hormone biosynthesis), DEPs and DCEs have more than two patterns, which indicates that members of DEPs and DCEs play different roles in different periods.

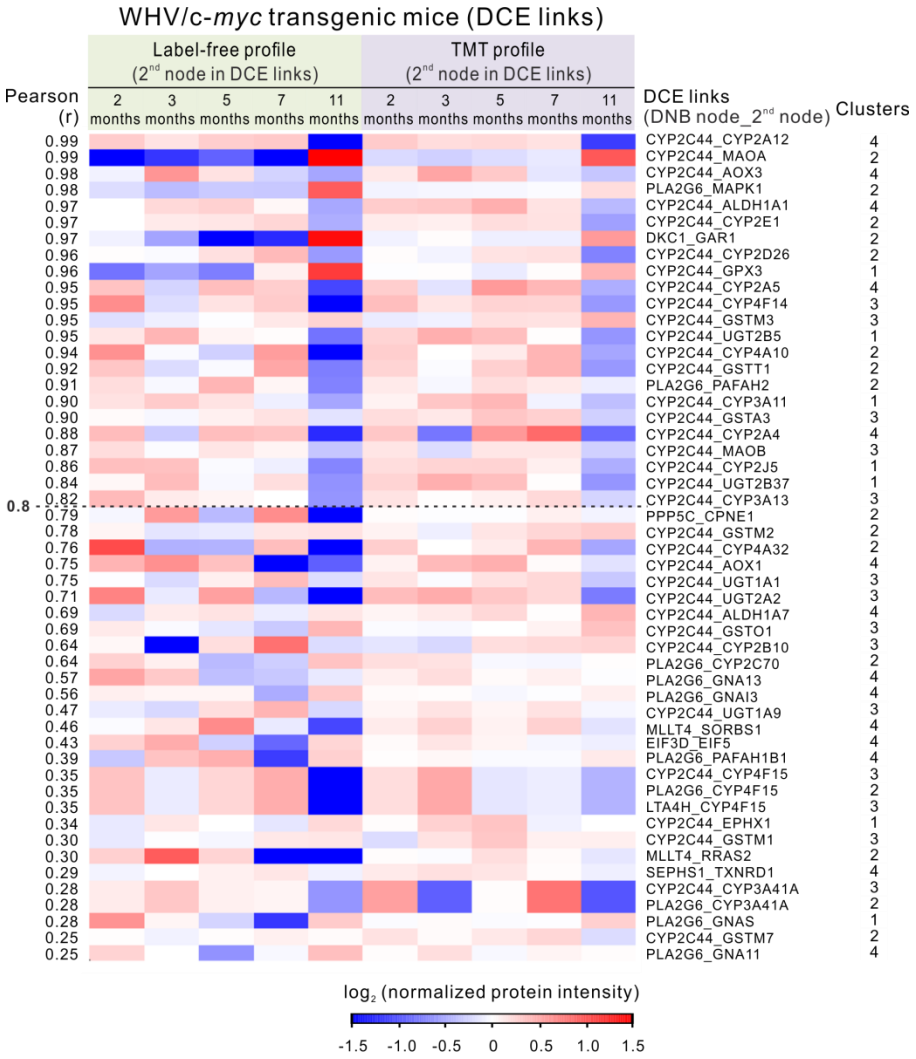

Figure S10
Label-free and TMT profiles of proteins of DNB-associated DCEs for WHV/c-*myc* transgenic mice.

A heatmap shows the reproduced dynamic profiles (label-free and TMT, individually) for the linked (second) nodes of DNB-associated DCEs mentioned in Figure 3. This experiment result validated DCEs in WHV/c-*myc* transgenic mice.

Supplementary References

1. Liu Y, Li C, Xing Z, Yuan X, Wu Y, Xu M, Tu K, Li Q, Wu C, Zhao M, Zeng R: **Proteomic mining in the dysplastic liver of WHV/c-myc mice--insights and indicators for early hepatocarcinogenesis.** *FEBS J* 2010, **277:**4039-4053.
2. Wisniewski JR, Zougman A, Nagaraj N, Mann M: **Universal sample preparation method for proteome analysis.** *Nat Methods* 2009, **6:**359-362.
3. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28:**27-30.
4. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M: **Biogrid: A General Repository for Interaction Datasets.** *Nucleic Acids Res* 2006, **34:**D535-539.
5. Zhao F, Xuan Z, Liu L, Zhang MQ: **TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies.** *Nucleic Acids Res* 2005, **33:**D103-107.
6. Chuang HY, Lee E, Liu YT, Lee D, Ideker T: **Network-based classification of breast cancer metastasis.** *Mol Syst Biol* 2007, **3:**140.
7. Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, et al: **A proteome-scale map of the human interactome network.** *Cell* 2014, **159:**1212-1226.
8. Wang J, Huang Q, Liu ZP, Wang Y, Wu LY, Chen L, Zhang XS: **NOA: a novel Network Ontology Analysis method.** *Nucleic Acids Res* 2011, **39:**e87.
9. Futschik M: **Mfuzz: Soft clustering of time series gene expression data.** 2015**:**R package version 2.30.30, http://www.sysbiolab.eu/software/R/Mfuzz/index.html.