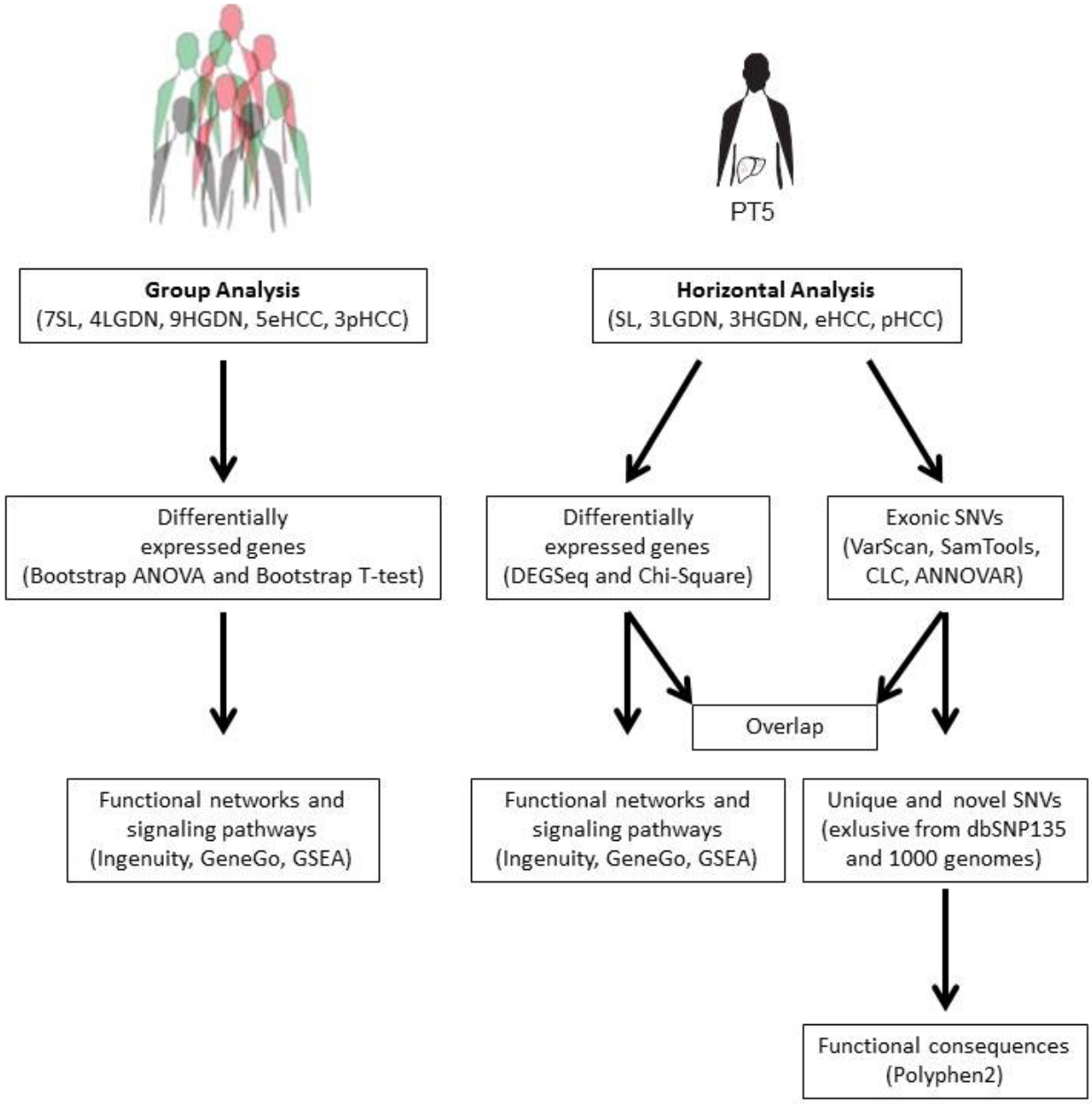
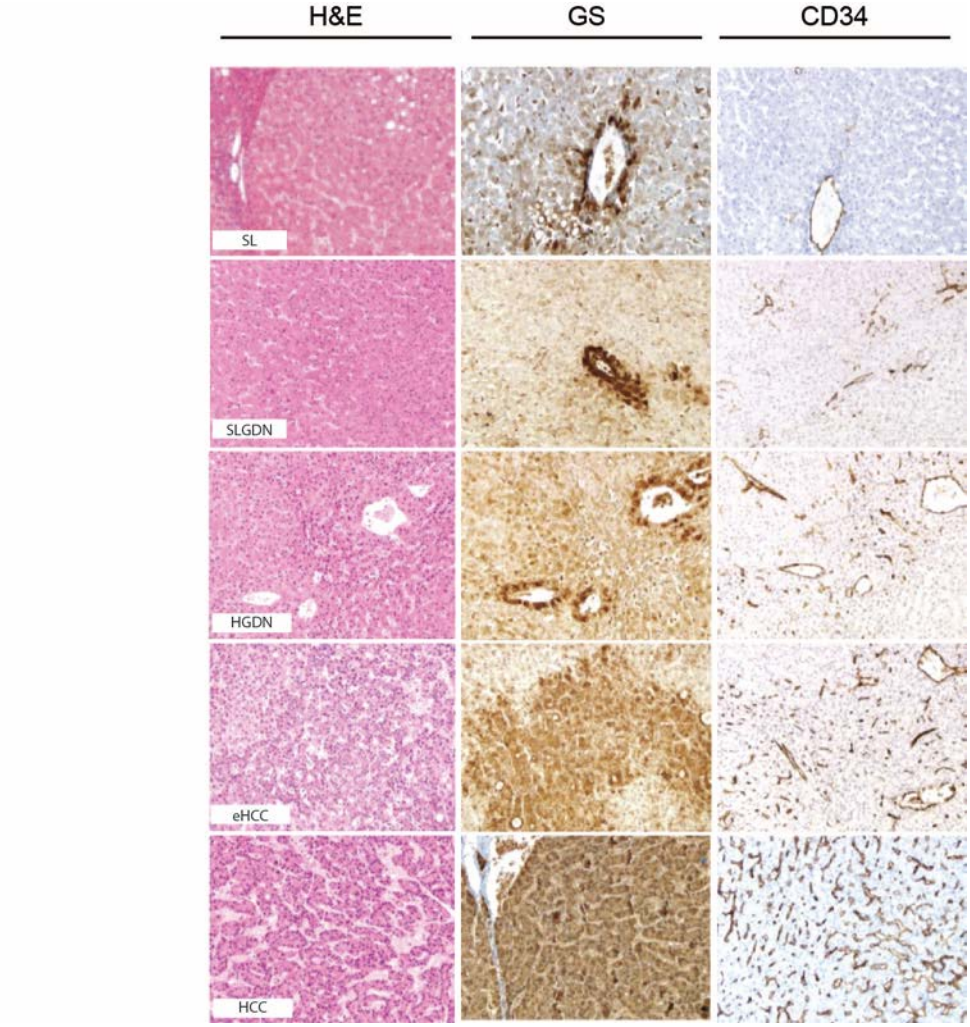
Supplementary Material



Supplementary Fig.1 Schematic overview of the analytical pipeline. Shown are the analyses of hepatic lesions from all patients (Group Analysis)) and horizontal analysis of lesions from the patient 5 (PT5) using the indicated algorithms. Exonic SNVs in PT5 were identified after exclusion of the SNVs recognized in dbSNP135 and 1000 genomes.



Supplementary Fig. 2 Microscopic and immunohistochemical features of a cirrhotic nodule and surrounding liver (SL), low grade dysplastic nodule (LGDN), high grade dysplastic nodule (HGDN), eHCC and progressed HCC in a 61 year old patient with B viral cirrhosis (patient 5). Shown is a gradual increase in expression of glutamine synthetase (GS) and neoangiogenesis marker CD34 with progression of hepatocarcinogenesis. Original magnification, X 200.

A **GeneGo analysis**

Dysplastic lesions: 234 genes

No.	Pathway Maps	p-Value	Ratio
1	2-Naphthylamine and 2-Nitronaphtalene metabolism	5.62E-07	7/61
2	Androstenedione and testosterone biosynthesis and metabolism p.2	1.78E-04	4/35
3	Glutathione metabolism	1.80E-04	5/65
4	Glutathione metabolism / Human version	1.94E-04	5/66
5	Androstenedione and testosterone biosynthesis and metabolism p.2/ Rodent version	1.99E-04	4/36
6	Glutathione metabolism / Rodent version	2.73E-04	5/71
7	NRF2 regulation of oxidative stress response	9.58E-04	4/54
8	Development_Role of CDK5 in neuronal development	2.62E-03	3/34
9	Apoptosis and survival_APRIL and BAFF signaling	3.89E-03	3/39
10	Apoptosis and survival_Anti-apoptotic TNFs/NF-kB/Bcl-2 pathway	4.48E-03	3/41

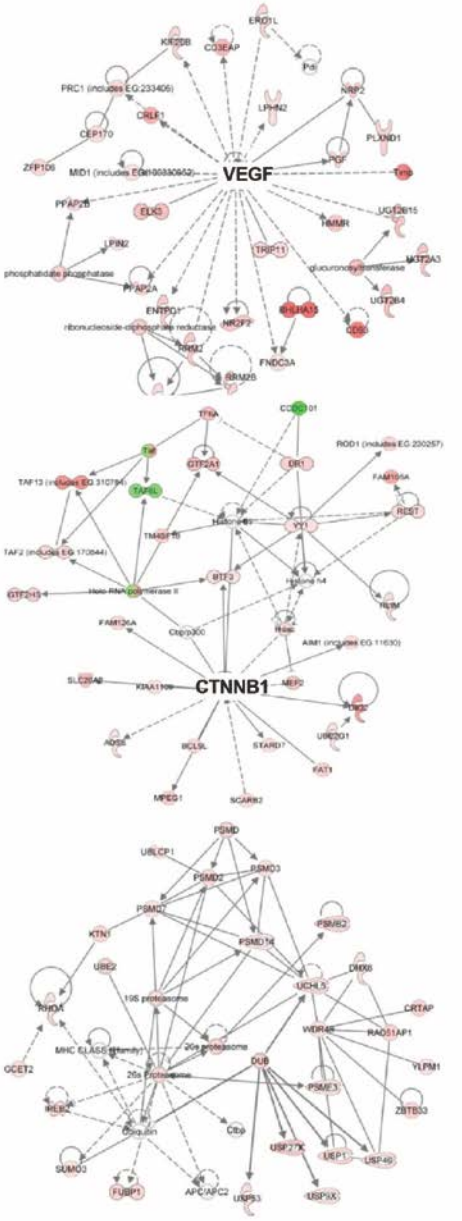
Dysplastic to eHCC: 123 genes

No.	Pathway Maps	p-Value	Ratio
1	Arachidonic acid production	6.67E-04	3/50
2	Stellate cells activation and liver fibrosis	1.85E-03	3/71
3	Cell cycle_Role of 14-3-3 proteins in cell cycle regulation	2.59E-03	2/22
4	Immune response_IFN alpha/beta signaling pathway	3.08E-03	2/24
5	Cell cycle_Initiation of mitosis	3.34E-03	2/25
6	G-protein signaling_Ras family GTPases in kinase cascades (scheme)	3.61E-03	2/26
7	Immune response_CD137 signaling in immune cell	4.48E-03	2/29
8	Immune response_ETV3 affect on CSF1 -promoted macrophage differentiation	5.10E-03	2/31
9	Cell cycle_Role of APC in cell cycle regulation	5.43E-03	2/32
10	Cell cycle_Spindle assembly and chromosome separation	5.77E-03	2/33

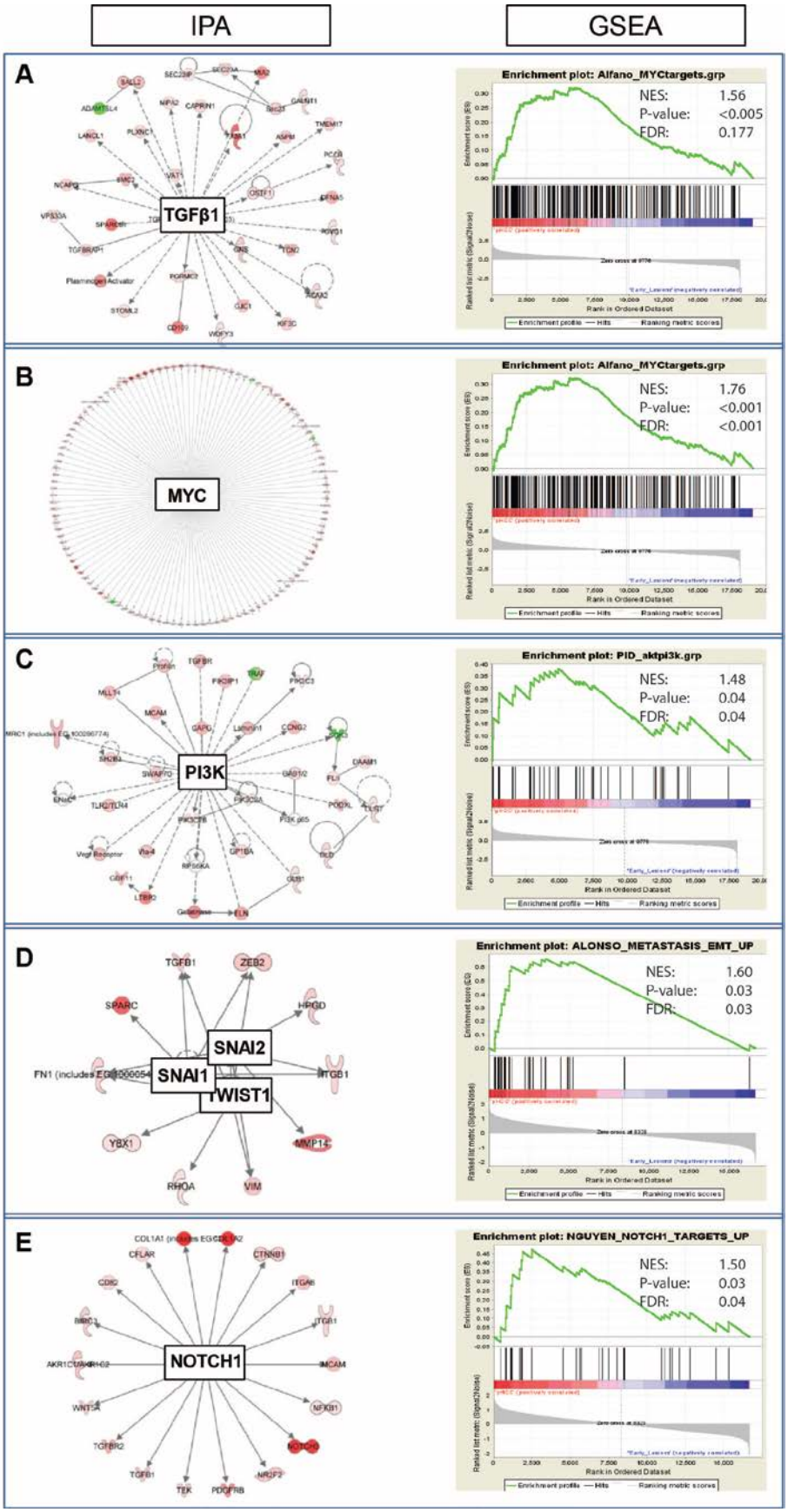
Progressed HCC -specific: 1486 genes

No.	Pathway Maps	p-Value	Ratio
1	Cytoskeleton remodeling_Cytoskeleton remodeling	2.15E-12	29/102
2	Cell adhesion_Chemokines and adhesion	7.83E-12	28/100
3	Cell adhesion_ECM remodeling	1.00E-09	18/52
4	Cytoskeleton remodeling_TGF, WNT and cytoskeletal remodeling	3.11E-09	26/111
5	Cell adhesion_Integrin-mediated cell adhesion and migration	1.21E-07	15/48
6	Cell adhesion_Integrin inside-out signaling	1.86E-07	16/56
7	Development_TGF-beta-dependent induction of EMT via MAPK	6.21E-07	14/47
8	Gamma-secretase proteolytic targets	7.11E-07	18/76
9	Stellate cells activation and liver fibrosis	1.23E-06	17/71
10	Development_Regulation of epithelial-to-mesenchymal transition (EMT)	1.36E-06	16/64
11	Cell cycle_Spindle assembly and chromosome separation	2.94E-06	11/33
12	Development_TGF-beta-dependent induction of EMT via SMADs	5.65E-06	11/35
13	Cytoskeleton remodeling_Integrin outside-in signaling	6.59E-06	13/49
14	Cell adhesion_Endothelial cell contacts by non-junctional mechanisms	7.91E-06	9/24
15	Immune response_IL-7 signaling in B lymphocytes	8.45E-06	12/43

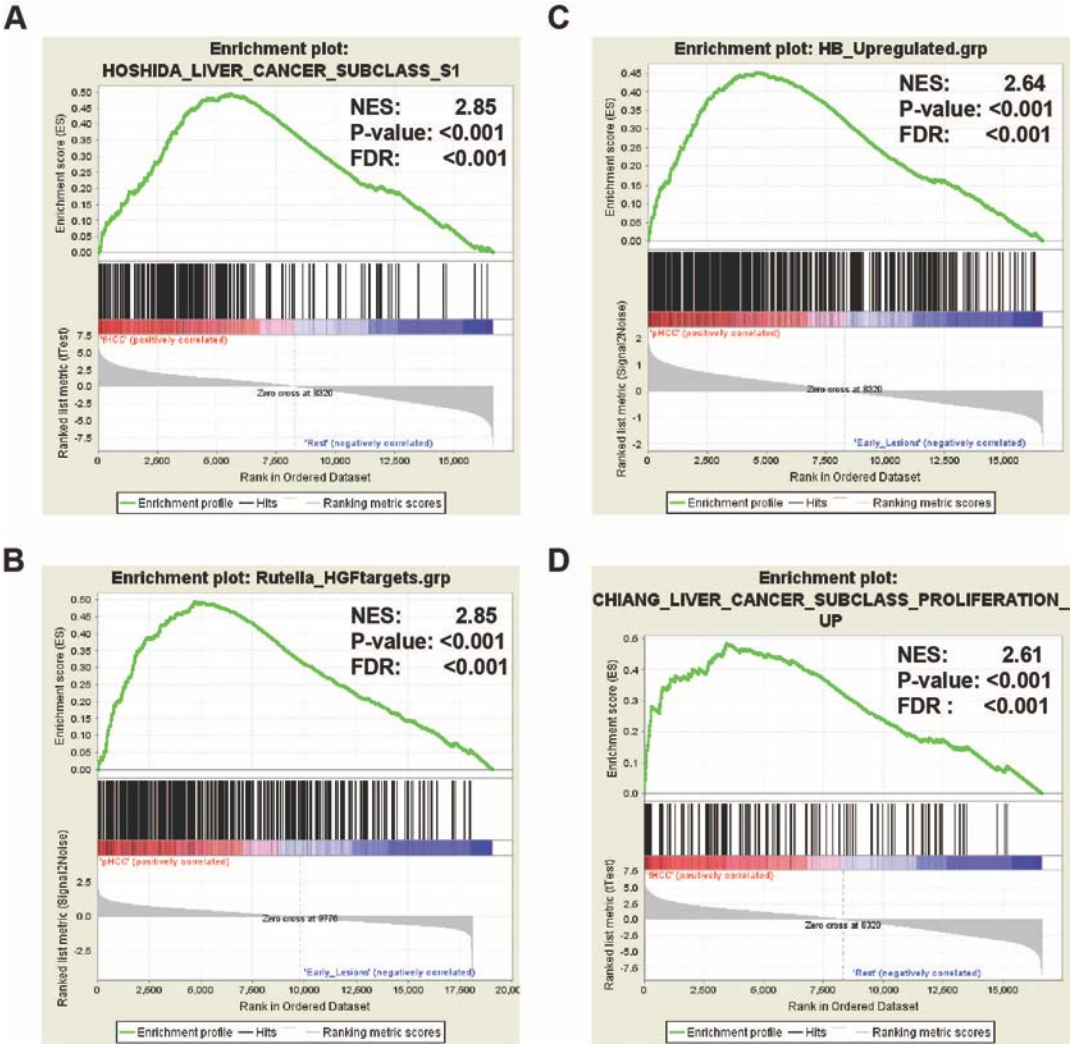
B **IPA**



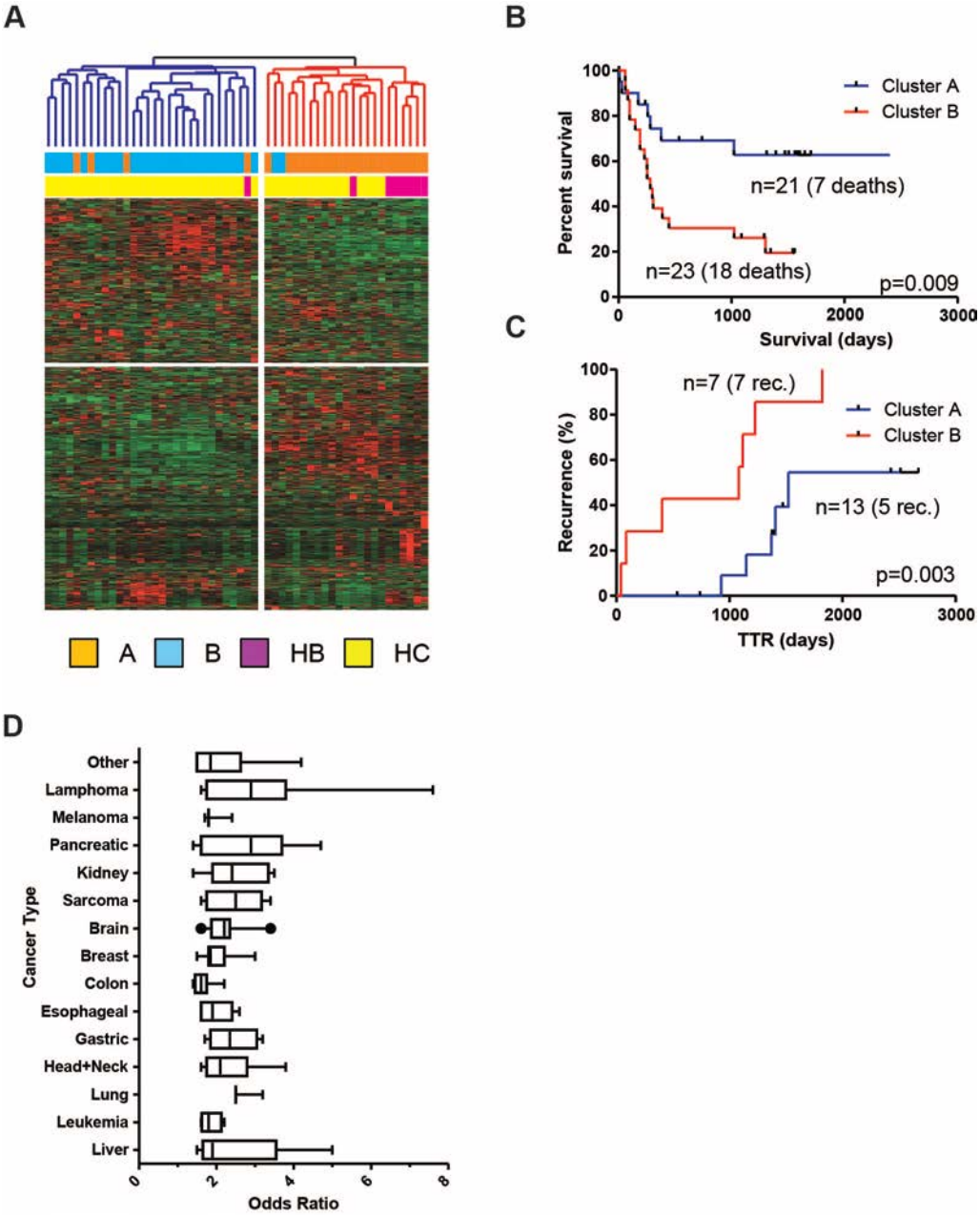
Supplementary Fig. 3 Transcriptomic changes during evolution of HCC. (A) GeneGo analysis of the dominant functional pathways at the sequential stages of hepatocarcinogenesis. (B) IPA analysis of dominant signaling pathways identified in pHCC including VEGF (top), β-Catenin (middle) and protein degradation (bottom).


IPA
GSEA
A
TGFβ1
Enrichment plot: Alfano_MYCtargets.grp
NES: 1.56
P-value: <0.005
FDR: 0.177
B
MYC
Enrichment plot: Alfano_MYCtargets.grp
NES: 1.76
P-value: <0.001
FDR: <0.001
C
PI3K
Enrichment plot: PID_akt pi3k.grp
NES: 1.48
P-value: 0.04
FDR: 0.04
D
TGFB1
ZEB2
SPARC
HPGD
SNAI2
SNAI1
ITGB1
TWIST1
YBX1
MMP14
RHOA
VIM
Enrichment plot: ALONSO_METASTASIS_EMT_UP
NES: 1.60
P-value: 0.03
FDR: 0.03
E
CFLAR
CTNNB1
ITGAV
ITGB1
NOTCH1
MCAM
NFKB1
TGFBR2
TGFB1
TEK
PDGFRB
NR2F2
Enrichment plot: NGUYEN_NOTCH1_TARGETS_UP
NES: 1.50
P-value: 0.03
FDR: 0.04

Supplementary Fig. 4 Activation of key functional networks and prognostically relevant gene sets associated with poor HCC prognosis identified by IPA and GSEA and centered on (A) TGFβ [16]; (B) MYC [13]; (C) PI3K [21]; (D) SNAI1 and SNAI2 [14]; and (E) NOTCH1[19]. Normalized Enrichment score (*NES)* reflects degree of over-representation for each group at the peak of the entire set. Statistical significance calculated by nominal *P*-value of the *NES* by using an empirical phenotype-based permutation test. False positives are calculated by the false discovery rate (FDR).

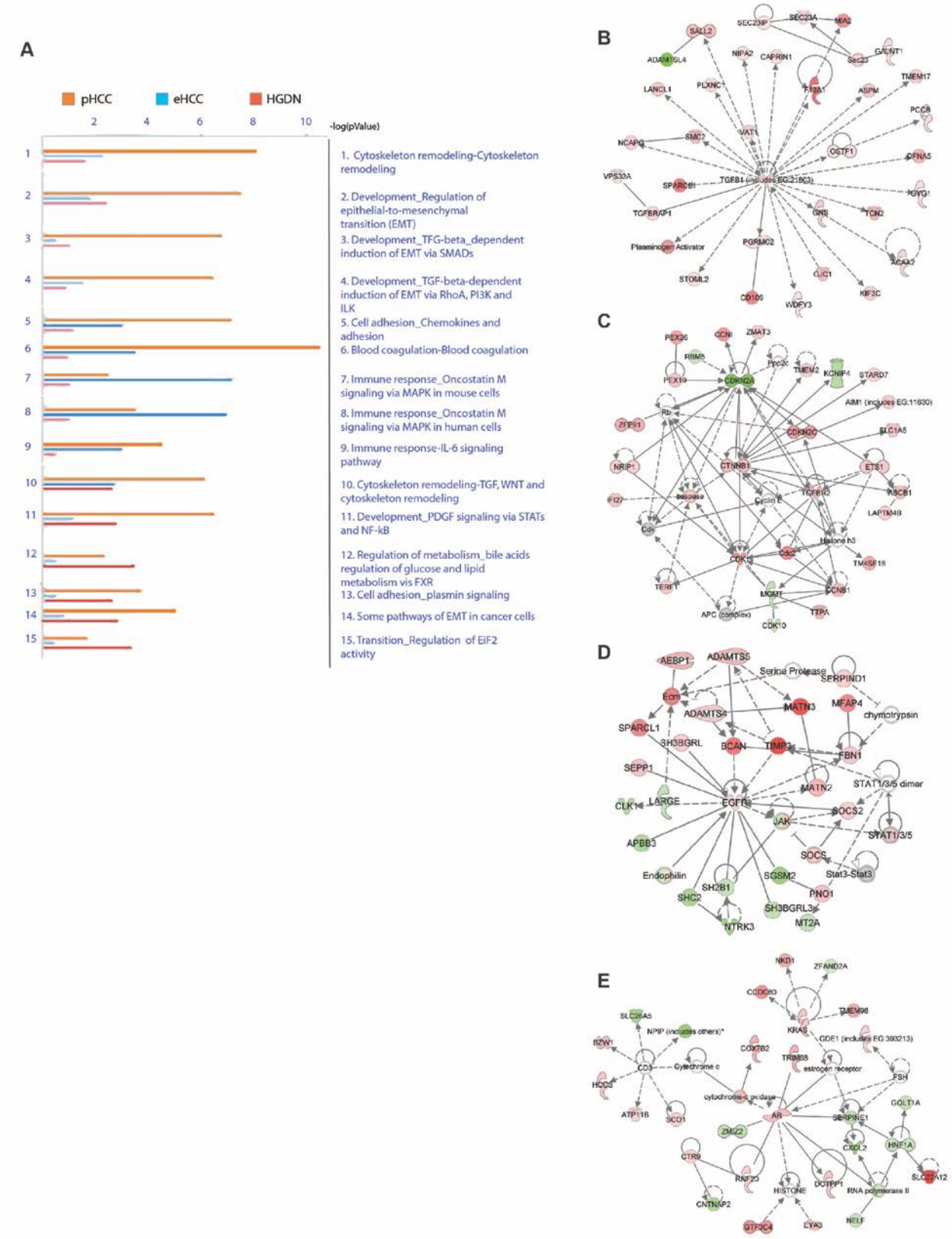


Supplementary Fig. 5 GSEA analysis of prognostically relevant gene sets in pHCC associated with poor survival. (A) WNT [17]; (B) HGF/MET [20]; (C) hepatoblastoma subtype [18]; and (D) proliferation [15]. Normalized enrichment score (*NES*) reflects degree of over-representation for each group at the peak of the entire set. Statistical significance calculated by nominal *P*-value of the *NES* by using an empirical phenotype-based permutation test. False positives are calculated by the false discovery rate (FDR).

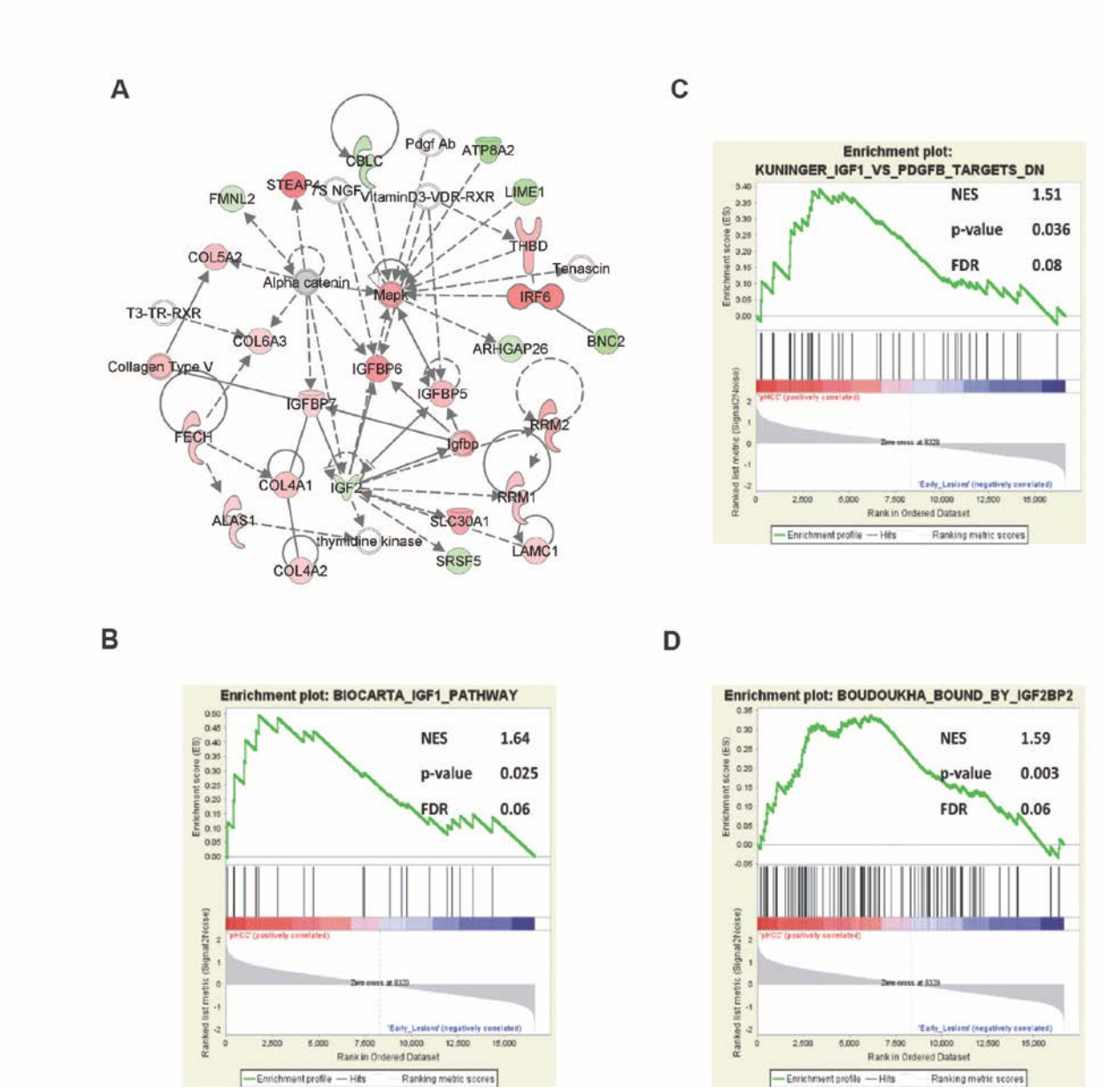


Supplementary Fig. 6 Prognostic significance of the pHCC signature for HCC and other types of cancer. (A) Hierarchical cluster analysis of 53 human HCCs based on pHCC 1486-gene signature. Bars under cluster tree represent overlap with previously generated HCC subtypes A, B, HB and HC described by Lee *et al*.[18]. (B,C) Kaplan-Meier plots and log-rank statistics of overall survival (B) and recurrence (C) of the HCC patients (Mantel-Cox test). Recurrence data were available only for 20 patients. (D) Association of the pHCC signature with clinical outcome of patients with different types of cancer. Integrative meta-analysis of genomic data using the

Oncomine Microarray database. Data shown as means odds ratio ± SD with *P*-values < 0.001 and odds ratios >2 set as a threshold.


A
pHCC
eHCC
HGDN
2
4
6
8
10
-log(pValue)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
1. Cytoskeleton remodeling-Cytoskeleton remodeling
2. Development_Regulation of epithelial-to-mesenchymal transition (EMT)
3. Development_TFG-beta_dependent induction of EMT via SMADs
4. Development_TGF-beta-dependent induction of EMT via RhoA, PI3K and ILK
5. Cell adhesion_Chemokines and adhesion
6. Blood coagulation-Blood coagulation
7. Immune response_Oncostatin M signaling via MAPK in mouse cells
8. Immune response_Oncostatin M signaling via MAPK in human cells
9. Immune response-IL-6 signaling pathway
10. Cytoskeleton remodeling-TGF, WNT and cytoskeleton remodeling
11. Development_PDGF signaling via STATs and NF-kB
12. Regulation of metabolism_bile acids regulation of glucose and lipid metabolism vis FXR
13. Cell adhesion_plasmin signaling
14. Some pathways of EMT in cancer cells
15. Transition_Regulation of EiF2 activity
B
SEC23IP
SEC23A
MIA2
SALL2
ADAMTSL4
NIPA2
CAPRIN1
Sec23
GALNT1
TMEM17
LANCL1
PLXNC1
ASPM
PCCB
NCAPG
SMC2
VAT1
OSTF1
DPYSL5
VPS33A
SPARCL1
TGFB1 (includes EG:21803)
HGVG1
TGFBRAP1
GNS
TCN2
Plasminogen Activator
PGRMC2
ACAA2
STOML2
CUC1
CD109
KIF3C
WDFY3
C
PEX26
CCNI
ZMAT3
RBM5
Ppp2c
PEX19
CDKN2A
TMEM2
KCNIP4
STARD7
AIM1 (includes EG:11630)
Rb
ZFP91
CDKN2C
SLC1A5
NRIP1
CTNNB1
ETS1
IFI27
Cyclin E
TGFBR2
ABCB1
LAPTM4B
Cdk
Histone h3
Cdc25
CDK1
TM4SF18
TERF1
MGMT
CCNB1
APC (complex)
TTPA
CDK10
D
AEBP1
ADAMTS5
Serine Protease
SERPIND1
Ecm
ADAMTS4
MATN3
MFAP4
chymotrypsin
SPARCL1
SH3BGRL
BCAN
TIMP3
FBN1
SEPP1
MATN2
STAT1/3/5 dimer
CLK1
LARGE
EGFR
SOCS2
JAK
STAT1/3/5
APBB3
SOCS
Endophilin
SGSM2
Stat3-Stat3
SH2B1
PNO1
SHC2
SH3BGRL3
NTRK3
MT2A
E
NKD1
ZFAND2A
CCDC80
SLC26A5
TMEM98
NPIP (includes others)*
KRAS
GDE1 (includes EG:393213)
BZW1
COX7B2
TRIM38
CD3
Cytochrome c
estrogen receptor
FSH
HCCS
cytochrome-c oxidase
GOLT1A
ATP11B
SCO1
AR
SERPINE1
ZMIZ2
HNF1A
CXCL2
CTR9
RNF20
DCTPP1
SLC27A12
HISTONE
RNA polymerase II
CNTNAP2
GTF3C4
LYAR
NELF

Supplementary Fig. 7 Functional networks and pathways enriched in pHCC from patient 5. (A) GeneGo analysis of activated pathways in the different types of lesions. (B-E) IPA analysis of the major signaling networks centered on (B) TGFβ, (C) β-Catenin (CTNNB1); (D) EGFR and (E) androgene receptor (AR).



Supplementary Fig. 8 Disruption of IGF signaling by the loss of IGFALS in pHCC. (A) Functional networks in pHCC centered on IGF signaling determined by IPA analysis. (B-D) GSEA analysis of gene sets involved in IGF signaling: (B) Biocarta, (C) IGF-induced versus PDGF-induced targets [42] and (D) targets bound by IGFBP2 [41]. Normalized enrichment score (*NES*) reflects degree of over-representation for each group (define better group) at the peak of the entire set. Statistical significance calculated by nominal *P*-value of the *NES* by using an empirical phenotype-based permutation test. False positives are calculated by the false discovery rate (FDR).