



ARTICLE

DOI: 10.1038/s41467-017-00991-w

OPEN

# Integrative analysis of genomic and epigenomic regulation of the transcriptome in liver cancer

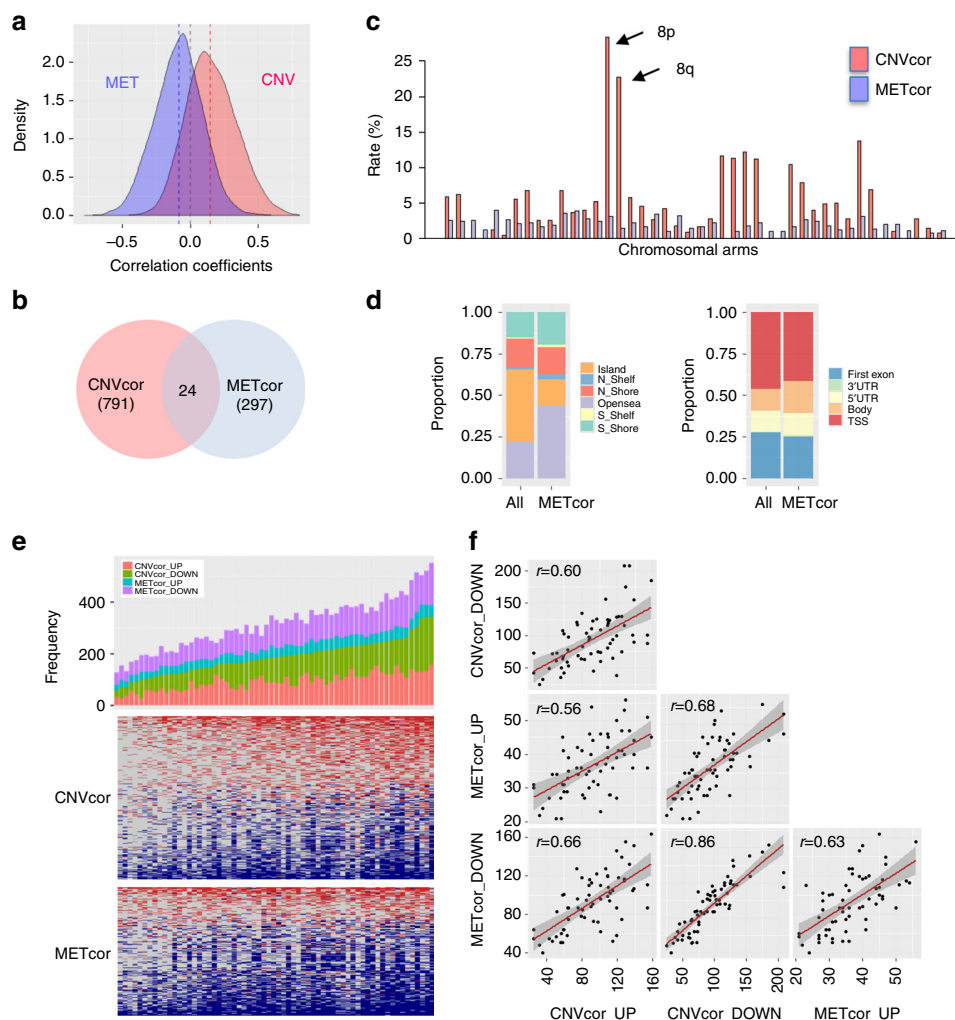
Hyun Goo Woo <sup>1,2</sup>, Ji-Hye Choi<sup>1,2</sup>, Sarah Yoon<sup>1</sup>, Byul A. Jee<sup>1,2</sup>, Eun Ju Cho<sup>3</sup>, Jeong-Hoon Lee <sup>3</sup>, Su Jong Yu<sup>3</sup>, Jung-Hwan Yoon<sup>3</sup>, Nam-Joon Yi<sup>4</sup>, Kwang-Woong Lee<sup>4</sup>, Kyung-Suk Suh<sup>4</sup> & Yoon Jun Kim<sup>3</sup>

Hepatocellular carcinoma harbors numerous genomic and epigenomic aberrations of DNA copy numbers and DNA methylation. Transcriptomic deregulation by these aberrations plays key driver roles in heterogeneous progression of cancers. Here, we profile DNA copy numbers, DNA methylation, and messenger RNA expression levels from 64 cases of hepatocellular carcinoma specimens. We find that the frequencies of the aberrancies of the DNA copy-number-correlated (CNVcor) expression genes and the methylation-correlated expression (METcor) genes are co-regulated significantly. Multi-omics integration of the CNVcor and METcor genes reveal three prognostic subtypes of hepatocellular carcinoma, which can be validated by an independent data. The most aggressive subtype expressing stemness genes has frequent *BAP1* mutations, implying its pivotal role in the aggressive tumor progression. In conclusion, our integrative analysis of genomic and epigenomic regulation provides new insights on the multi-layered pathobiology of hepatocellular carcinoma, which might be helpful in developing precision management for hepatocellular carcinoma patients.

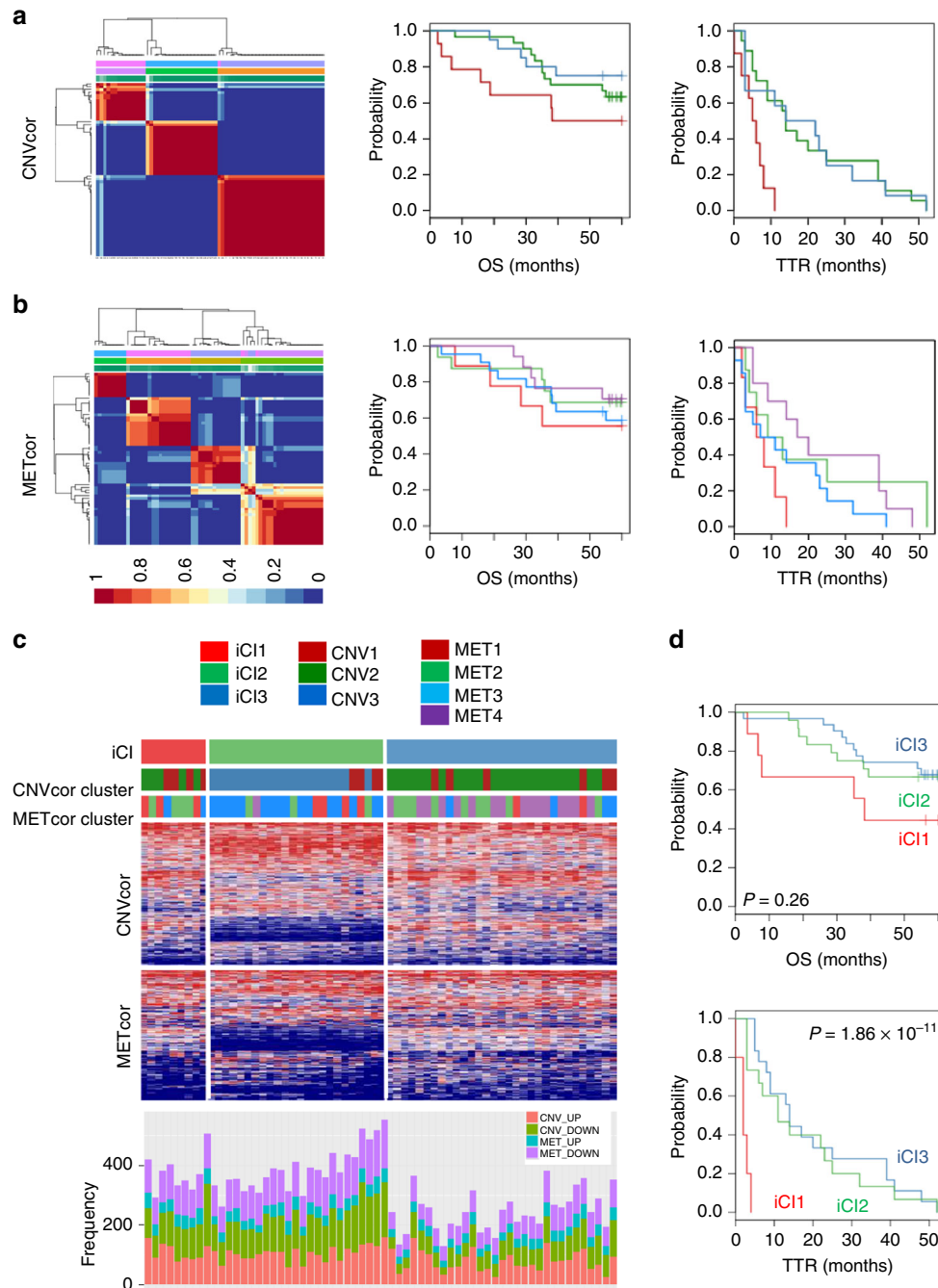
<sup>1</sup>Department of Physiology, Ajou University School of Medicine, 164 Worldcup-ro, Yeongtong-gu, Suwon 16499, Republic of Korea. <sup>2</sup>Department of Biomedical Sciences, Graduate School, Ajou University, 164 Worldcup-ro, Yeongtong-gu, Suwon 16499, Republic of Korea. <sup>3</sup>Department of Internal Medicine, Liver Research Institute, Seoul National University College of Medicine, 28 Yongon-dong, Chongno-gu, Seoul 110-744, Republic of Korea. <sup>4</sup>Department of Surgery, Seoul National University College of Medicine, 28 Yongon-dong, Chongno-gu, Seoul 110-744, Republic of Korea. Correspondence and requests for materials should be addressed to Y.J.K. (email: [yoonyun@snu.ac.kr](mailto:yoonyun@snu.ac.kr))

Recent large-scale and multi-omics profiling of cancers has provided a systematic picture of genomic and epigenomic deregulation in these diseases. Genomic alterations due to DNA copy-number aberration or mutations occur frequently during tumorigenesis, stimulating cancer progression. Epigenetic regulation of the cancer genome by DNA methylation also plays pivotal roles in heterogeneous cancer behaviors. In particular, in hepatocellular carcinoma (HCC), genomic profiling studies have demonstrated the enormous heterogeneity of genomic and epigenomic deregulation<sup>1</sup>. In this cancer, aberrations of DNA copy number play key regulatory roles in HCC progression<sup>2–4</sup>, and transcriptional deregulation resulting from such aberrations is a potential driver event in HCC progression<sup>5, 6</sup>. In addition, DNA methylation profiling studies have revealed the biological and clinical significance of epigenetic regulation in HCC progression<sup>7–11</sup>. Several key cancer-related genes such as *IGF2*<sup>12</sup> and *UHRF1*<sup>13</sup> exert their regulatory functions by modulating DNA methylation.

However, despite the genome-wide impact of aberrations of DNA copy numbers and DNA methylation on cancers, it remains unclear whether DNA copy-number aberration is systematically related to epigenetic DNA methylations, and, if so, whether this connection plays any role in cancer progression. In this study, we profiled DNA copy numbers, DNA methylation, and messenger RNA (mRNA) expression levels in a cohort of HCC patients. To identify genes whose expression levels are regulated by genomic and/or epigenomic deregulation, we defined DNA copy-number-correlated (CNVcor) and DNA methylation-correlated (METcor) genes, based on their corresponding gene expression levels across samples, respectively. CNVcor genes represent the transcriptional deregulation dependent on DNA copy number, whereas METcor genes represent transcriptional deregulation dependent on DNA methylation. Expression of CNVcor genes was significantly correlated with expression of METcor genes, suggesting concomitant regulation



**Fig. 1** Identification of DNA copy-number-correlated (CNVcor) and DNA methylation-correlated (METcor) genes in HCC. **a** Distribution of the correlation coefficients between the mRNA expression levels and DNA copy numbers or DNA methylation across the samples are shown, respectively. **b** A Venn diagram shows the counts of CNVcor and METcor genes. The counts of overlapped genes between CNVcor and METcor are indicated. **c** Proportional frequencies of the CNVcor or METcor genes against total gene counts in each chromosome arm are shown, respectively. **d** Genomic positions of DNA methylation probes are categorized based on the positional relations with CpG islands (right) and genes (left), respectively. The proportional frequencies in each category of DNA methylation for the whole genes and METcor genes are compared, respectively. **e** A barplot shows the frequencies of up- or down-expressed CNVcor genes or METcor genes in each sample (top). Each of CNVcor and METcor genes is categorized as upregulated (CNVcor\_UP and METcor\_UP) or downregulated (CNVcor\_DOWN and METcor\_DOWN) genes, respectively. Heatmaps show the up- (red) and down-expressed (blue) CNVcor (middle) and METcor (bottom) genes, respectively. **f** Dot plots show the pairwise correlations among the frequencies of CNVcor\_UP, CNVcor\_DOWN, METcor\_UP, and METcor\_DOWN genes, respectively



**Fig. 2** Identification of molecular subtypes of HCC using CNVcor and METcor genes. **a, b** Plots show the non-negative factorization (NMF) cluster results for the CNVcor in CNV data (**a**) and for the METcor in MET data (**b**), respectively. Kaplan–Meier plot analyses for subtypes identified by NMF clustering of the CNVcor and METcor genes are shown for overall survival (OS) and time to tumor recurrence (TTR), respectively. **c** Heatmaps show the expression patterns of subtypes identified by iCluster analysis. The subtypes identified by the CNVcor or METcor genes using NMF cluster methods are indicated with colored bars (top). The frequencies of aberrant expression of CNVcor\_UP, CNVcor\_DOWN, METcor\_UP, and METcor\_DOWN genes in each subtype are shown (bottom). **d** Kaplan–Meier plot analyses for subtypes identified by iCluster (iCI1, iCI2, and iCI3) are shown for overall survival (OS) and time to tumor recurrence (TTR), respectively

of cancer transcriptomes by alterations in genomic DNA copy number in addition to epigenetic DNA methylation. Moreover, by performing multi-omics integration of CNVcor and METcor genes, we could identify distinct molecular subtypes that were significantly associated with prognostic outcomes of HCC. Further systematic analysis could identify new mutations that could be used as precision targets or biomarkers for subtype distinction.

**Results**

**Transcriptome deregulation by DNA copy number or methylation.** Genomic and epigenomic profiles of DNA copy-number variation (CNV), DNA methylation (MET), and gene expression (EXP) were obtained from 64 HCC patients. Raw data profiles were preprocessed as described in “Methods.” To assess the global effects of genomic and/or epigenomic aberrations, we calculated the correlation coefficients between DNA copy number or DNA

**Table 1 Clinico-pathological features of the molecular subtypes of SNU data**

	iC11 (n = 9)	iC12 (n = 24)	iC13 (n = 31)	P value ( $\chi^2$ -test)
<b>Sex</b>				
Male	8	19	29	0.276
Female	1	5	2	
<b>Age</b>				
<60 (years)	5	12	16	0.960
≥60 (years)	4	12	15	
<b>Stage</b>				
T1/T2	5	21	27	0.065
T3/T4	4	3	4	
<b>Nodal metastasis</b>				
No	7	24	30	0.060
Yes	1	0	0	
<b>Tumor size</b>				
<5 cm	6	10	12	0.319
≥5 cm	3	14	19	
<b>Vascular invasion</b>				
No	1	15	19	<sup>a</sup> 0.002
Yes	7	16	5	
<b>Gross type</b>				
Simple nodular	3	16	14	0.110
Multi nodular/ infiltrative	6	7	16	
<b>AFP</b>				
<400 (ng/ml)	5	18	25	0.286
≥400 (ng/ml)	4	5	6	
<b>Grade</b>				
I, II	2	11	19	0.278
III, IV	4	5	10	

<sup>a</sup>Statistical significance  $P < 0.05$ 

methylation profiles with the corresponding mRNA expression profiles. The correlation coefficient  $r$  was normalized to stabilize variance by applying Fisher's Z-transformation:  $z = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right)$ . Consistent with previous studies, the overall distribution of correlation coefficients between DNA copy numbers and the corresponding gene expression profiles exhibited a significant skew to the right (skewness = 0.224,  $P < 2.2 \times 10^{-16}$ , D'Agostino test). By contrast, the correlation coefficients between DNA methylation profiles and corresponding gene expression profiles were skewed to the left (skewness = -0.153,  $P = 7.9 \times 10^{-11}$ ) (Fig. 1a). This result reflects the overall impacts of positive and negative transcriptional deregulation due to aberrant DNA copy numbers and DNA methylation, respectively.

We next identified the positively correlated gene signatures for DNA copy number (CNVcor,  $n = 813$ ) and the negatively correlated gene signatures for DNA methylation (METcor,  $n = 321$ ), which were determined based on the Fisher's Z-transformed correlation coefficients with 95% confidence interval ( $\geq 1.96$  or  $\leq -1.96$ ,  $P < 0.05$ ). (Supplementary Data 1). CNVcor genes exhibited DNA copy-number-dependent transcriptional deregulation, whereas METcor genes exhibited DNA methylation-dependent transcriptional deregulation. The CNVcor and METcor were not likely to overlap each other showing only 24 genes were overlapped, which may imply exclusive regulation of CNVcor and METcor genes in transcriptional deregulation (Fig. 1b). The CNVcor genes exhibited a regional genomic preference for DNA copy-number aberration,

particularly on the  $p$  and  $q$  arms of chromosome 8 (Fig. 1c). Consistent with previous studies, the abundance of the CNVcor genes on chromosome 8 indicated regional sensitivity of gene expression to DNA dosage<sup>5, 14</sup>. By contrast, the METcor genes were not distributed in preferred chromosome regions. Moreover, in comparison with all probes, METcor genes frequently resided in the inter-genic open sea regions rather than CpG islands (Fig. 1d, left). In relation to genetic position, DNA methylation was more frequent in gene body regions than near transcription start site (TSS) or first-exon regions (Fig. 1d, right), indicating that methylation in open sea and/or gene body regions might function in transcriptional deregulation. In addition, gene ontology analysis revealed differential enrichment of gene functions between CNVcor and METcor genes. CNVcor genes were enriched in functions related to protein transport (enrichment score, ES = 4.1), metabolism (ES = 3.88), and the cell cycle (ES = 2.29), whereas METcor genes were enriched in functions related to inflammation (ES = 3.41) and metabolism (ES = 3.12) (Supplementary Fig. 1). Although further studies are required, our results imply that CNVcor and METcor genes play distinct functional roles in transcriptional deregulation of HCC.

Because the CNVcor and METcor genes were thought as the aberrations with concomitant transcriptional deregulation, we further interrogated the association between the expression of the CNVcor and METcor genes by calculating the frequencies of transcriptional deregulation of both types of genes in each patient. This analysis revealed that patients with frequent deregulation of the CNVcor genes were more likely to exhibit frequent deregulation of the METcor genes (Fig. 1e). In addition, we designated up and downregulated CNVcor (CNVcor\_UP and CNVcor\_DOWN) and METcor (METcor\_UP and METcor\_DOWN) genes whose expression levels differed by more than twofold between tumors and surrounding non-tumoral tissues. Pairwise comparison of the frequencies of up- and down-regulated CNVcor or METcor genes revealed significant correlations between the two sets ( $P < 0.001$ , Pearson's correlation test, Fig. 1f). Thus, we suggest that the frequencies of DNA copy-number- and DNA methylation-dependent transcriptional deregulation were highly correlated with each other.

**Molecular subtypes based on CNVcor and METcor genes.** Next, we investigated whether the expression profiles of CNVcor and METcor genes could predict prognostic subgroups. For each gene set profile, we applied a non-negative matrix factorization (NMF) cluster analysis with cluster number  $k$  from 2 to 5, and then determined  $k$  for each profile: CNV,  $k = 3$ ; MET,  $k = 4$  (Fig. 2 and Supplementary Fig. 2). Interestingly, the subtypes identified by CNVcor overlapped significantly with the subtypes identified by METcor ( $P = 1.36 \times 10^{-5}$ ,  $\chi^2$ -test), consistent with the associated regulation of CNVcor and METcor genes in HCC. In addition, Kaplan–Meier plot analyses demonstrated that the subtypes identified by CNVcor or METcor genes could predict patients' clinical outcomes of time to recurrence (TTR), respectively (CNVcor: TTR,  $P = 0.004$ ; METcor: TTR,  $P = 0.07$ ). Kaplan–Meier plot of overall survival (OS) also revealed distinct prognostic outcomes among the groups, although the difference was not statistically significant (CNVcor: OS,  $P = 0.22$ ; METcor: OS,  $P = 0.7$ ), possibly due to the smaller sample size in each subtype.

We next sought to identify molecular subtypes that reflected the multi-layered expression patterns of the CNVcor and METcor genes. To integrate the genomic data regarding DNA copy numbers, DNA methylation, and RNA expression, we employed an integrated cluster method, iCluster<sup>15</sup>. Clustering analysis was performed with cluster numbers  $k = 3$  and 4, which were determined as the number of clusters from the NMF analyses on

the individual data sets, respectively. This analysis revealed three distinct subtypes with  $k=3$ , iC1, iC2, and iC3, which were consistent with the classification based on the NMF analysis for individual CNV and MET data, respectively ( $P=3.65 \times 10^{-11}$  and  $1.09 \times 10^{-5}$ ;  $\chi^2$ -test). Interestingly, the frequencies of aberrant CNVcor and METcor genes were much higher in subgroups iC1 and iC2 than in iC3 (Fig. 2c, bottom), indicating that the subtypes were distinct from one other concerning the correlated regulation of transcriptomes by genomic and epigenomic aberrations.

Notably, Kaplan–Meier plot analysis revealed that, among the three subgroups, iC1 had the worst prognostic outcomes of TTR and OS (TTR,  $P=1.86 \times 10^{-11}$ ; OS,  $P=0.26$ , Fig. 2d). We also compared clinico-pathological features among the iC1, iC2, and iC3 subgroups, and found that iC1 tumors had more frequent vascular invasion than iC2 or iC3 tumors ( $P=0.002$ , Table 1). Other clinical features did not differ among the subgroups. These results are consistent with the distinct aggressiveness properties of the molecular subtypes. Based on these findings, we suggest that integrated analysis of CNVcor and METcor genes can identify molecular subtypes, each of which features a different combination of genomic and epigenomic features related to transcriptional deregulation, associated with distinct prognostic outcomes.

**Validation of the molecular subtypes in an TCGA data set.** To validate the robustness and consistency of our molecular classification using the CNVcor and METcor genes, we applied the same method to an independent HCC data set from TCGA (LIHC, Liver Hepatocellular Carcinoma). Consistent with the aforementioned analysis, in the LIHC data, we observed that the overall distribution of correlation coefficients between the CNV and transcriptome was shifted to the right (skewness = 0.367,  $P < 2.2 \times 10^{-16}$ , D’Agostino test), whereas the correlation coefficients for DNA methylation were shifted to the left (skewness = -0.237,  $P < 2.2 \times 10^{-16}$ ) (Fig. 3a). CNVcor ( $n=388$ ) and METcor ( $n=1496$ ) genes were identified from the LIHC data using the same analysis pipeline (Fig. 3b). Consistent with the Seoul National University (SNU) data, CNVcor genes were enriched on chromosome 8, whereas the METcor genes did not exhibit a chromosomal preference (Fig. 3c). METcor genes were more frequent in open sea and gene body regions than in CpG islands and TSS regions (Fig. 3d). In addition, the correlation between the frequencies of aberrant CNVcor and METcor genes was validated in the LIHC data ( $r=0.82$ ,  $P=1.69 \times 10^{-88}$ , Fig. 3e).

In addition, we determined the molecular subtypes of LIHC data using the same method that we applied to the SNU data. This analysis also revealed three subtypes (C1, C2, and C3), which had clear differences in their frequencies of CNV and methylation aberrations (Fig. 3f). Group C1 had the highest rate of aberrations in DNA copy number and DNA methylation, whereas group C3 had the lowest rate (Fig. 3f, g). As expected, the three groups had distinct prognostic outcomes, with the worst OS ( $P=0.001$ ) and recurrence-free survival (RFS,  $P=0.015$ ) in group C1, and the most favorable clinical outcomes in the group C3 (Fig. 3h). These results confirmed that our strategy for classifying tumors based on CNVcor and METcor can identify prognostic subgroups with distinct genomic and epigenomic regulation that reflect the clinical outcomes of HCC patients. Thus, the profile of CNVcor and METcor genes can predict prognostic molecular subtypes independent of patient cohort and data platforms.

**Coordinated aberrations of DNA copy numbers and methylation.** We also compared the frequencies of the aberrant DNA copy number and DNA methylation in the entire genome in the integrated SNU and LIHC data set ( $n=428$ ) after correcting for batch

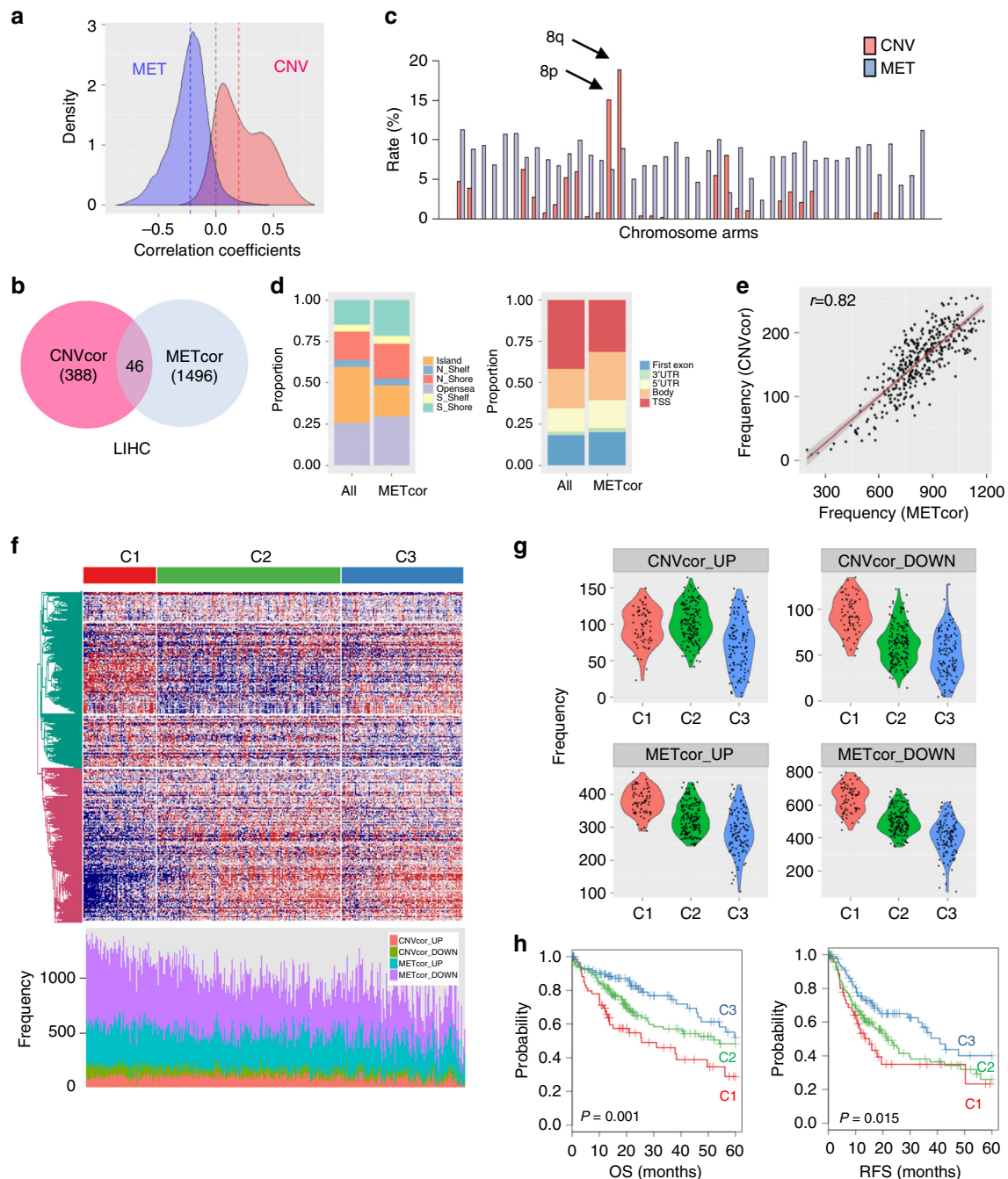
effects as described in “Methods.” The DNA copy-number gain (CNVgain) and loss (CNVloss) and the DNA hypermethylation (METHyper) and hypomethylation (METHypo) were determined with the cutoff of the fold difference 0.2 compared to the average value for each probe of non-tumoral tissues. The frequencies of CNVgain were significantly correlated with the frequencies of CNVloss ( $r=0.43$ ,  $P=1.5 \times 10^{-20}$ , Fig. 4a). By contrast, the frequencies of DNA hypermethylation were negatively correlated with the frequencies of DNA hypomethylation ( $r=-0.27$ ,  $P=2.25 \times 10^{-8}$ , Fig. 4b). In addition, as with the frequencies of CNVcor and METcor aberrations, the frequencies of the aberrant DNA copy number, including CNVgain and CNVloss, were significantly correlated with the frequencies of aberrant DNA methylation, including METHyper and METHypo ( $r=0.41$ ,  $P=2.25 \times 10^{-18}$ , Fig. 4c). Directional aberrations of CNVgain and CNVloss, and METHyper and METHypo, were also closely correlated, indicating that the correlations were independent of directional changes (Fig. 4d–g). Taken together, we suggest that the HCC patients with frequent aberration of DNA copy number are more likely to have frequent aberration of DNA methylation. These correlated frequencies of aberrant CNVcor and METcor genes may imply the close relationships between the aberrations of DNA copy number and DNA methylation, and which were not dependent to patient cohorts or data platforms.

#### Identification of genomic key features in the HCC subtypes.

Next, we examined the mutation profiles of LIHC to determine whether they were associated with our subclassification. After filtering out the synonymous mutations, we obtained gene-level mutations of missense and nonsense variants. Overall, mutation frequencies did not differ among the subtypes (Supplementary Fig. 3). Among the 189 recurrently mutated genes with more than 10 mutations across the samples, we identified 37 differentially mutated genes, which had >5% difference of the mutational frequencies among the subgroups of C1, C2, and C3 (Fig. 5a). Interestingly, *BAP1* was the most frequently mutated gene in C1 ( $n=10$ , 14.7%), but it was mutated less frequently in C2 ( $n=4$ , 2.25%) and C3 ( $n=6$ , 5.12%). *BAP1* expression is associated with multiple tumor types and high tumor phenotype penetrance<sup>16</sup>. By contrast, *CTNNB1* mutation was more frequent in the favorable prognostic groups (C2,  $n=18$ , 10.16%; C3,  $n=13$ , 11.11%), but less frequent in the aggressive subtype C1 ( $n=2$ , 2.91%). Supporting this, previous studies have shown that the *CTNNB1* mutation is associated with favorable HCC prognosis<sup>17</sup>. Taken together, these findings suggest that molecular HCC subtypes associated with DNA copy numbers and DNA methylation are also linked to *BAP1* and *CTNNB1* mutations, which might play regulatory roles in subtype progression of HCC.

In addition, to address the functional determinants of the subtypes, we performed a network analysis using the core CNVcor genes ( $n=95$ ) and METcor genes ( $n=179$ ) shared between the SNU and LIHC data. Interestingly, the *UBC* gene was highly connected by physical and genetic interactions to most of the core CNVcor genes (77 out of 95, 81%), although *UBC* itself was not a member of CNVcor genes (Fig. 5b). *UBC* encodes ubiquitin C, which regulates multiple biological functions, and its expression is associated with tumor progression<sup>18</sup>. This suggests that dysregulation of CNVcor genes could be effectively suppressed by modulating ubiquitination, although further studies would be required to investigate this possibility in detail. The core METcor genes did not have a significant hub gene in this network analysis.

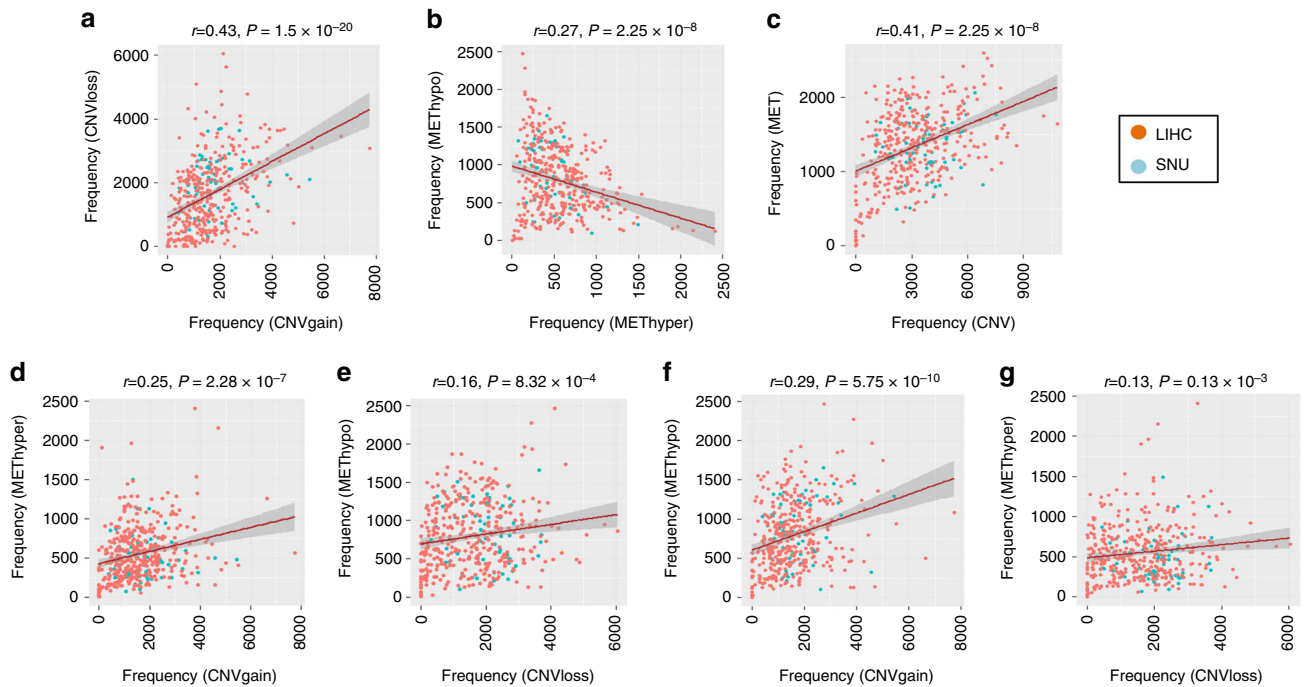
Furthermore, to delineate key distinctive features of the subtype transcriptomes, we identified differentially expressed genes (DEGs) among subtypes that are commonly found in



**Fig. 3** Validation of molecular subtypes using an independent TCGA data. **a** Overall distributions of correlation coefficients of between the mRNA expression levels and DNA copy numbers or DNA methylation are shown, respectively. **b** A Venn diagram shows the counts of the DNA copy number-correlated genes (CNVcor) and the DNA methylation-correlated genes (METcor) genes. The counts of the overlapped genes between CNVcor and METcor are indicated. **c** Proportion of the CNVcor or METcor genes against total count of genes in each chromosome arm are shown. **d** Composition of DNA methylation probes in the whole genes and METcor genes are compared. Genomic positions of DNA methylation probes are categorized based on the relations with CpG island regions (right) and gene regions (left), respectively. **e** A plot shows the correlation between the frequencies of CNVcor and METcor genes in each sample. **f** A heatmap shows the expression patterns of the differentially expressed genes among the subtypes identified by iClusterPlus analysis (top). The frequencies of CNVcor\_UP, CNVcor\_DOWN, METcor\_UP, and METcor\_DOWN genes are shown (bottom). **g** Violin plots indicate the frequencies of CNVcor\_UP, CNVcor\_DOWN, METcor\_UP, and METcor\_DOWN genes are shown in each subtype of C1, C2, and C3, respectively. **h** Kaplan-Meier plot analyses of the HCC subtypes for overall survival (OS) and recurrence-free survival (RFS) are shown, respectively

both SNU and LIHC data sets, including up- ( $n=121$ ) and down-expressed ( $n=366$ ) genes. Notably, we identified that *CA9* was the top-ranked gene that was differentially expressed in iC11/C1 tumors in comparison with other tumor subtypes (Fig. 5c, top). Indeed, *CA9* is a marker of hypoxia, and its overexpression is a poor prognostic marker in HCC<sup>19</sup>. In addition, in comparison

with the other subtypes, the aggressive iC11 and C1 tumors expressed high levels of stemness-related genes such as *KRT19*, *EPCAM*, and *PROM1*. Consistent with this, a recent study reported that *CA9* expression is associated with stemness-related phenotypes in HCC<sup>20</sup>. These results indicate that the aggressiveness of iC11 and C1 tumors might be associated with expression



**Fig. 4** Coordinated aberrations of DNA copy number and DNA methylation in liver cancer. **a, b** The SNU and LIHC data sets of DNA copy numbers or DNA methylation are integrated by applying “combat” method as described in “Methods.” Aberrations of DNA copy numbers or DNA methylation were determined with cutoff fold difference  $>0.2$  compared to those of the average values of the non-tumoral tissues, respectively. Directional alterations of DNA copy-number gain (CNVgain) and loss (CNVloss) and DNA hypermethylation (METHyper) and hypomethylation (METHypo) in each sample are plotted, respectively. **c** Overall frequencies of CNV aberration including CNVgain and CNVloss and DNA methylation including METHyper and METHypo are plotted. **d-g** Plots show the pairwise frequencies of CNVgain, CNVloss, METHyper, and METHypo genes in individual samples

of stemness traits. Based on this finding, we also evaluated the known stemness gene sets (i.e., gene sets for ESC, Nanog, Oct4, Sox2, Myc1, and Myc2). This analysis revealed that the iC1 and C1 tumors exhibited significant differential expression of stemness genes in comparison with the other subgroups (Fig. 5c, middle). CA9 expression might represent a promising candidate marker for the most aggressive subtype of tumors (iC1 or C1).

We also compared our classification with the previous studies that have defined molecular classifiers for HCC subtypes. By applying nine different molecular classifiers as described in “Methods,” we evaluated their enriched expression in SNU and LIHC data sets, respectively. When we compared each of the HCC classifier with the DEGs ( $n = 487$ ), only a subset of the HCC signature genes were overlapped with DEGs, suggesting that each classifier represents distinct biological characteristics (Fig. 5c, bottom left). However, Kim\_65 genes (34 out of 65), Yamashita\_EpCam\_DOWN (7 out of 18), Woo\_CLHCC\_DOWN (203 out of 374), and Boyault\_G123\_DOWN (26 out of 50) genes showed substantial overlaps with the DEGs ( $>20\%$ ), indicating their expression similarity of the subtypes from those studies. Although not all the classifiers were matched to our classification, most of them exhibited differential enrichment among the subtypes. Particularly, C1 and iC1 tumors exhibited differential expression of the prognostic signatures compared to the other subgroups, including Hoshida et al., Kim\_65 genes, Roesler\_PT\_UP, Woo\_CLHCC, Yamashita\_EpCAM, and Boyault\_G123 signatures (Fig. 5c, bottom right). This also imply that the expression of these previous classifiers might be associated with the frequencies of the aberrations of CNVcor and METcor genes. Collectively, we suggest that the C1 and iC1 tumors might share functional features of those subtypes such as aggressiveness or stemness.

As shown in Fig. 5a, c, *BAP1* mutation is linked with the expression of stemness in C1 tumors. To clarify this association, we evaluated the effects of *BAP1* suppression on the expression of stemness genes in liver cancer cells. Knockdown of *BAP1* expression by transfecting the *BAP1*-targeting short hairpin RNAs (shRNAs) in Huh7 cells significantly upregulated the expression of stemness genes including *CA9*, *KRT19*, *EPCAM*, and *PROM1* (Fig. 5d). These results strongly support our finding that the *BAP1* mutation may contribute, at least in part, to the progression of an aggressive HCC subtype expressing stemness genes.

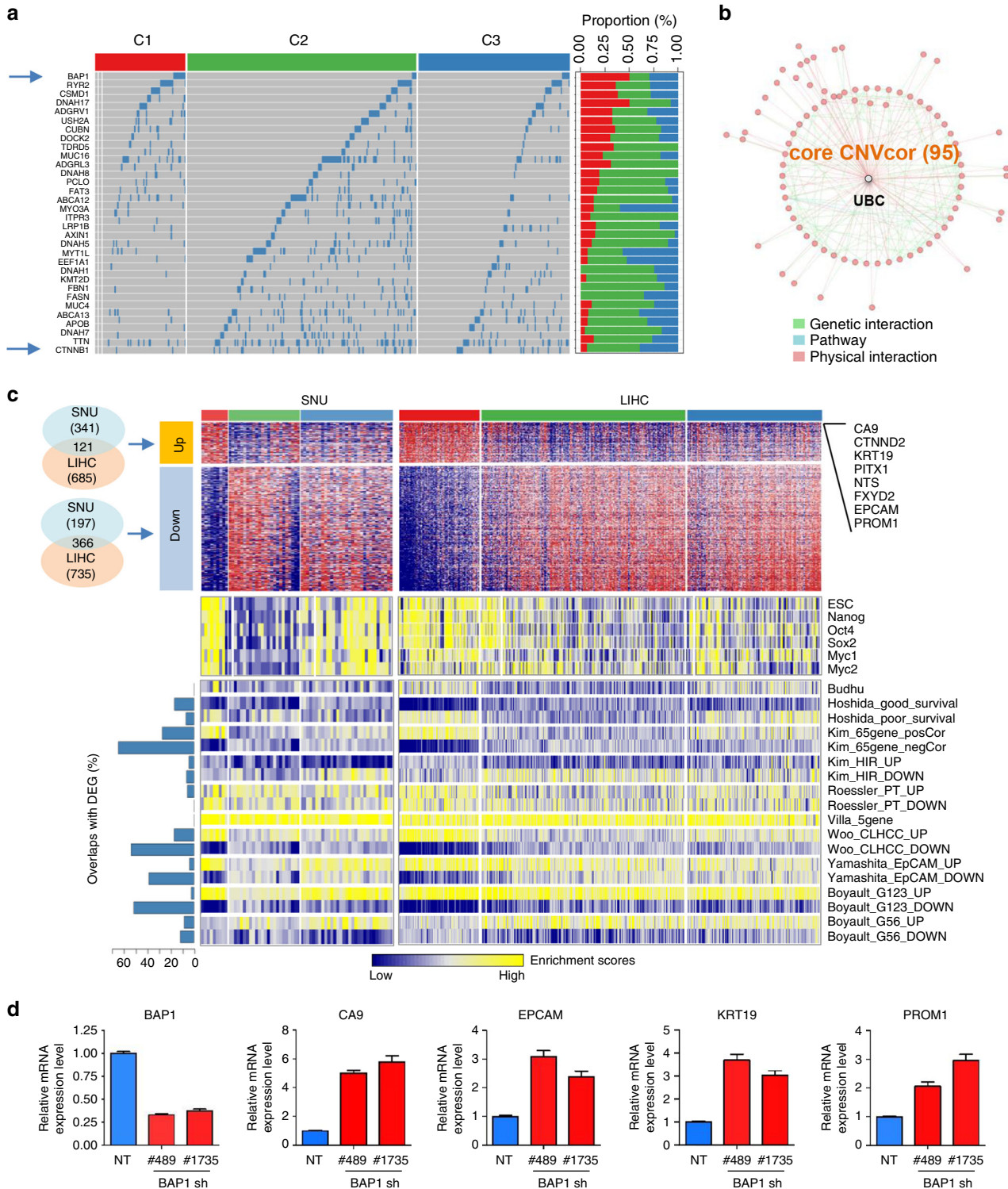
## Discussion

Previous studies demonstrated that integrative analysis of multi-layered genomic features of cancers could identify molecular subtypes, providing novel mechanistic and clinical insights into tumor heterogeneity, as well as revealed candidate therapeutic targets and biomarkers. However, unveiling the enormous complexity of cancer genome data remains challenging. Indeed, HCC has heterogeneous etiological backgrounds, which may also impede extraction of biologically meaningful information from genomic data. With this concern in mind, in this study, we enrolled patients who had a homogenous etiological background of hepatitis B virus (HBV) infection and had experienced the same clinical management in one hospital. As a strategy for integrating multi-layered genomic and epigenomic data, we defined the genomic and epigenomic deregulators of CNVcor and METcor genes, and demonstrated that these correlated genes could successfully identify subtypes of HCC that reflected their distinct, multi-layered molecular features, as well as their distinct prognostic outcomes. Our findings were validated in an independent TCGA data set, notwithstanding its heterogeneous sample composition. Moreover, we found that HCCs with

higher frequencies of CNVcor aberration had higher frequencies of METcor aberration, demonstrating that the patients who have frequent aberration of DNA copy numbers are prone to have frequent aberration of DNA methylation. This observation highlights the fact that aberrations of DNA copy number and methylation need to be considered together in data analysis.

Remarkably, our classification analysis based on CNVcor and METcor revealed novel molecular key features that represent new precision targets and biomarkers for HCC management. When we compared the mutation profiles among the subtypes, we

observed differential mutation rates of *BAP1* and *CTNNB1*. In particular, C1 tumors had frequent mutations at *BAP1* (14.7%), although less frequent in overall HCC samples. *BAP1*, a tumor suppressor gene, encodes a nuclear deubiquitinase that is thought to mediate its effects through chromatin modulation, transcriptional regulation, and possibly the ubiquitin-proteasome system and DNA damage response pathway. Recently, mutations in *BAP1* have been detected in various malignancies, in which they confer increased susceptibility for tumorigenesis<sup>21–24</sup>. Recurrent *BAP1* mutations have also been observed in intrahepatic





cholangiocarcinoma<sup>25</sup>, and a recent study showed that loss of *Bap1* in mice increases expression of H3K27me3 and Ezh2 while suppressing expression of *PRC2* targets<sup>26</sup>. Dysregulation of Polycomb genes is frequently observed in tumors expressing stemness traits<sup>27</sup>. This is consistent with our finding that C1 tumors had the top-ranked expression of stemness genes, including *CA9*, *KRT19*, *EPCAM*, and *PROM1* (encoding CD133) (Fig. 5c), well-known markers of stemness traits, as well as poorer prognostic outcomes in HCC<sup>28–31</sup>. Similarly, depletion of *BAP1* results in loss of differentiation and gain of stem-like properties in uveal melanoma cells<sup>32</sup>. In liver cancer cells, we also observed same results (Fig. 5d). These data raise a possibility that the aggressive behavior of C1 tumors might be mediated, at least in part, by *BAP1* mutation and the resultant induction of stemness traits.

In addition, by integrating our data with TCGA data, we identified *UBC* as a potential key regulator that acts as a hub gene for most CNVcor genes in the network. *UBC* encodes a component of the ubiquitin system whose activity yields free ubiquitin. The ubiquitin system has recently been identified as a target for cancer treatment<sup>18,33</sup>, and *UBC* is implicated in diverse tumor types, contributing to both tumorigenesis and cancer progression<sup>34,35</sup>. Thus, drugs targeting the ubiquitin/proteasome system, such as bortezomib, represent candidates for attenuating the aberrant regulation of CNVcor genes in aggressive iC1/C1 tumors.

Besides the most aggressive subtype iC1/C1, the favorable subtypes of iC2/C2 and iC3/C3 had no significant difference of prognostic outcomes. However, iC2/C2 and iC3/C3 subtypes showed distinct patterns on the levels of overall aberrations of DNA copy numbers and methylations (Fig. 2c and Fig. 3f, g). Further studies might be required to delineate underlying mechanisms for these discernible features of subtypes.

In conclusion, our systematic integration of genomic and epigenomic regulation of gene expression revealed coordinated multi-layered genomic aberrations of HCCs, and which could identify molecular subtypes of HCC providing novel mechanistic and clinical insights on the precision diagnostics and therapeutics for HCC patients.

## Methods

**Patients and tissue specimens.** A total of 64 cases HCC patients with HBV infection from SNU hospital were enrolled in this study. The study was approved by the Institutional Review Board of SNU Hospital (IRB number, 1211-063-442) and obtained written informed consents from donors.

### Profiling of DNA copy numbers, DNA methylation, and mRNA expression.

The frozen tissues from 64 cases of tumor specimens and 30 cases of the non-tumoral surrounding tissue specimens were used for genomic profiling. DNA copy-number variation (CNV), DNA methylation (MET), and mRNA expression (EXP) profiling in the same patients were performed using Human Omni1 chips, Infinium Human Methylation 27 BeadChip, and Human HT-12 Expression Bead-Chips as manufacturer's instruction, respectively (Illumina, San Diego, CA, USA).

Gene expression profile was normalized by log<sub>2</sub> transformation, quantile normalization, and aggregated by HUGO official symbols. Each data profile was normalized by subtracting the average values per probe of the non-tumoral tissues to represent the fold differences between the tumor and non-tumoral tissues. Then, for CNV data, gene-level DNA copy numbers in each sample were mapped with the segmented CNV values by circular binary segmentation algorithm implemented in R package library "DNAcopy." For DNA methylation profile, probe level  $\beta$ -values were filtered to remove the probes located on sex chromosomes. Then, the probes residing in CpG islands-related regions including CpG islands, Shelf, and Shore regions, differentially methylated regions, and gene promoter regions including the upstream 2500 bases from TSS, 5'UTR, and first-exon regions were mapped to their corresponding genes. For each of the processed profiles, the probes with >30 % of missing values across samples were removed, and the remained missing values were imputed by R package library "impute."

For each data set, genomic coordinates of the probes in each data set were updated to human reference genome hg38 by using R package library "liftOver," then the probes were matched to the corresponding probes of the EXP profiles. The tumor specific alteration was calculated by subtracting the average probe intensities of the non-tumoral tissues. After filtering out the probes with missing values >50% and the probes for sex chromosomes, data were imputed by *s* k-nearest neighbor algorithm. Then, pairwise Pearson's correlation coefficients were calculated for each gene in the paired profiles of CNV vs. EXP and MET vs. EXP, respectively. In cases where multiple probes mapped to a gene, the probe with a mean or a minimum value of the correlation coefficients was used as a representative pair-matched probe for the CNV and MET profiles, respectively.

**Clustering analysis of multi-layered genomic profiles.** NMF cluster analysis with standard "brunet" method and 50 iterations was employed to identify stable sample clusters using CNVcor and METcor genes, respectively. Number of clusters *k* was set to 2 to 5, and the preferred number of cluster was determined using the observed consensus map and cophenetic correlation between clusters, and the average silhouette width of the consensus membership matrix was determined by the R package "NMF." The minimum number of a cluster member was set to 10. For integrated cluster analysis for CNV, MET, and EXP profiles, we applied "iCluster" method implemented in R package with default parameters and 50 iterations<sup>15</sup>.

**Validation of the molecular subtypes using TCGA data.** Validation of the analysis results was performed using a data set of liver hepatocellular carcinoma (LIHC) from TCGA. Multi-layered profiles for DNA copy numbers, DNA methylation, mRNA expression, and mutations of LIHC data were obtained from official TCGA data portal (<https://tcga-data.nci.nih.gov>). After matching sample labels in each platform, we used 364 data sets that have the matched data sets of DNA copy numbers, methylation, and mRNA expression profiles. The data were preprocessed by applying same procedures used in our data set (SNU). For data integration of LIHC and SNU data, batch effects between the data sets were corrected using empirical Bayes methods implemented in an R Package "combat." For mutation data analysis, we used 362 samples matched to the integrated data set excluding the two samples that did not have mutation profiles (i.e., TCGA-G3-A25X-01 and TCGA-BC-4072-01).

**Gene set enrichment and network analyses.** For gene set analysis, the embryonic stem (ES) cell-related gene sets including ESC (ES cell), Nanog, Oct, Sox2, Myc1, and Myc2 were obtained from a previous study<sup>36</sup>. HCC classifiers from nine different studies were obtained at original publication sites, which included the gene signatures of Budhu et al.<sup>37</sup>, Villa et al.<sup>38</sup>, Hoshida\_survival<sup>39</sup>, Yamashita\_EpCam<sup>40</sup>, Kim\_65 gene<sup>41</sup>, Kim\_HIR (hepatic injury and regeneration)<sup>42</sup>, Boyault et al., Roessler\_PT (portal vein thrombosis)<sup>43</sup>, and Woo\_CLHCC (cholangiocarcinoma-like HCC)<sup>30</sup> (Supplementary Table 1). For each sample, the enrichment of a gene signature was calculated by applying Kolmogorov–Smirnov test. Briefly, for each individual gene expression profile, two *P* values for the

**Fig. 5** Identification of key molecular features for HCC subtypes. **a** Differentially mutated genes among the subtypes of LIHC are shown. A total of 37 differentially mutated genes are identified, which have differential mutation rates in C1 tumors compared to C2 or C3 tumors, respectively. The two samples of TCGA-G3-A25X-01 and TCGA-BC-4072-01 with no mutation data were excluded in the analysis (indicated with a right grey color). **b** Network of CNVcor genes is constructed using physical and genetic interactions from GeneMania software, which identifies the highest interconnecting hub gene, *UBC*. **c** Heatmaps show the differentially up- (DEG\_UP, *n* = 121) and down-expressed (DEG\_DOWN, *n* = 366) genes that are commonly found in both SNU and LIHC data sets, respectively (top). Top-ranked DEGs are indicated. The DEGs for SNU data are defined with fold difference >1.4 in comparison of iC1 tumors with iC2 or iC3 tumors, while the DEG for LIHC data are defined with fold difference >2 and permuted Student's *t* test *P* < 0.001 in comparison of the C1 tumors compared to C2 or C3 tumors, respectively. Gene set enrichment scores of the embryonic stem cell (ESC)-related gene sets (middle) and the nine molecular classifiers for HCC (bottom) are shown in SNU and LIHC data sets, respectively. Stemness gene sets of ESC, Nanog, Oct4, Sox2, Myc1, and Myc2, and the nine different HCC classifiers are obtained from previous studies as described in "Methods." **d** Huh7 cells stably expressing non-targeting (NT) shRNA and *BAP1* shRNA (#489 or #1735) were established. Real time qPCR for *BAP1* and stemness genes (*CA9*, *KRT19*, *EPCAM*, and *PROM1*) was performed. The expression values represent means  $\pm$  S.E.M. of three independent experiments

estimates  $D_+$  and  $D_-$  were calculated by Kolmogorov–Smirnov test, which determines the significance of the directional (positive or negative) enrichment of distributions of the signature. The ES  $S_{D_+}$  and  $S_{D_-}$  for a given signature was calculated as  $-\log_{10}(P \text{ value})$  from  $D_+$  and  $D_-$ , respectively. The ES  $S$  was defined as  $S_{D_+}$  if  $S_{D_+} > S_{D_-}$  and  $-S_{D_-}$  if  $S_{D_+} < S_{D_-}$ . In addition, canonical gene ontology analyses for biological processes were performed using DAVID software (<https://david.ncifcrf.gov>). Gene networks for given gene signatures were constructed using pathway, physical, and genetic interactions that were obtained from GeneMania plugin implemented in Cytoscape software (<http://www.cytoscape.org>).

**Cell culture and lentiviral vector transfection.** The Huh7 HCC cell line was obtained from the Korea Cell Line Bank (KCLB; Seoul, Korea). 293TN cells were obtained from System Biosciences (Mountain View, CA, USA), and were cultured in DMEM supplemented with 10% FBS, 100 U/ml penicillin, and 100  $\mu\text{g}/\text{ml}$  streptomycin. Lentiviral constructs expressing non-target (NT) and BAP1 shRNAs were purchased from Sigma-Aldrich (SHCLNG-NM\_004656; St. Louis, MO, USA), and transfected into 293TN cells (System Biosciences) with Lipofectamin 3000 transfection reagent (Invitrogen, Waltham, MA, USA). Two reaction tubes with 500  $\mu\text{l}$  aliquots of opti-MEM (Invitrogen; A and B) were prepared. Tube A was added with 24  $\mu\text{l}$  P3000 reagent and 8  $\mu\text{g}$  of DNA (3  $\mu\text{g}$  pGag-pol, 1  $\mu\text{g}$  VSV-G, and 4  $\mu\text{g}$  target plasmid). Tube B was added with 15  $\mu\text{l}$  lipofectamin 3000. Then, the mixture of tube A to tube B was incubated for 5 min and added in small drops to the cells with 70–90% confluence. The particles were collected 2 days after transfection. The lentivirus-infected cells were puromycin-selected for 1 week and stabilized by culturing for 4 weeks.

**Real-time PCR.** Cells were collected and total RNA was isolated using an RNeasy kit (Qiagen, Venlo, the Netherlands). The PrimeScript RT kit (Takara, Shiga, Japan) was used to reverse-transcribe the mRNA into complementary DNA. PCR was done using a CFX96 Real Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) with Ssoadvanced Universal Supermixes (Bio-Rad). The sequences of primers were described in Supplementary Table 2. Analysis of each sample was performed at least three times for each experiment, and the relative expression levels were measured by average values of  $2^{-\Delta\Delta\text{CT}} \pm \text{S.D.}$

**Statistical analysis.** All the statistical analysis was performed using the R package (<http://www.r-project.org>). OS was defined as the time from surgery to death. TTR was defined as the time from surgery date to the date of any diagnosed relapse. The follow-up time OS and TTR was truncated to 5 years.

**Data availability.** The raw data of the genomic profiles are available in GEO database (<http://www.ncbi.nlm.nih.gov/projects/geo>) with an accession number GSE87630.

Received: 20 February 2017 Accepted: 10 August 2017

Published online: 10 October 2017

## References

- Zucman-Rossi, J., Villanueva, A., Nault, J.-C. & Llovet, J. M. Genetic landscape and biomarkers of hepatocellular carcinoma. *Gastroenterology* **149**, 1226–1239.e4 (2015).
- Guichard, C. et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genetics* **44**, 694–698 (2012).
- Wang, K. et al. Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. *Hepatology* **58**, 706–717 (2013).
- Roessler, S. et al. Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival. *Gastroenterology* **142**, 957–966.e12 (2012).
- Woo, H. G. et al. Identification of potential driver genes in human liver carcinoma by genomewide screening. *Cancer Res.* **69**, 4059–4066 (2009).
- Kwon, S. M. et al. Genomic copy number alterations with transcriptional deregulation at 6p identify an aggressive HCC phenotype. *Carcinogenesis* **34**, 1543–1550 (2013).
- Stefanska, B. et al. Definition of the landscape of promoter DNA hypomethylation in liver cancer. *Cancer Res.* **71**, 5891–5903 (2011).
- Shen, J. et al. Genome-wide DNA methylation profiles in hepatocellular carcinoma. *Hepatology* **55**, 1799–1808 (2012).
- Villanueva, A. et al. DNA methylation-based prognosis and epdrivers in hepatocellular carcinoma. *Hepatology* **61**, 1945–1956 (2015).
- Revill, K. et al. Genome-wide methylation analysis and epigenetic unmasking identify tumor suppressor genes in hepatocellular carcinoma. *Gastroenterology*, **145**, 1424–35 (2013).
- Shen, J. et al. Exploring genome-wide DNA methylation profiles altered in hepatocellular carcinoma using Infinium HumanMethylation 450 BeadChips. *Epigenetics* **8**, 34–43 (2013).
- Martinez-Quetglas, I. et al. IGF2 is upregulated by epigenetic mechanisms in hepatocellular carcinomas and is an actionable oncogene product in experimental models. *Gastroenterology*, **151**, 1192–1205 (2016).
- Mudbhary, R. et al. UHRF1 overexpression drives DNA hypomethylation and hepatocellular carcinoma. *Cancer Cell* **25**, 196–209 (2014).
- Fehrmann, R. S. et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
- Shen, R., Olshen, A. B. & Ladanyi, M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* **25**, 2906–2912 (2009).
- Carbone, M. et al. BAP1 and cancer. *Nat. Rev. Cancer* **13**, 153–159 (2013).
- Hsu, H. C. et al. Beta-catenin mutations are associated with a subset of low-stage hepatocellular carcinoma negative for hepatitis B virus and with favorable prognosis. *Am. J. Pathol.* **157**, 763–770 (2000).
- Tang, Y. et al. Downregulation of ubiquitin inhibits the proliferation and radioresistance of non-small cell lung cancer cells in vitro and in vivo. *Sci. Rep.* **5**, 9476 (2015).
- Huang, W. J. et al. Expression of hypoxic marker carbonic anhydrase IX predicts poor prognosis in resectable hepatocellular carcinoma. *PLoS ONE* **10**, e0119181 (2015).
- Rhee, H. et al. Poor outcome of hepatocellular carcinoma with stemness marker under hypoxia: resistance to transarterial chemoembolization. *Mod. Pathol.* **29**, 1038–1049 (2016).
- Harbour, J. W. et al. Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science* **330**, 1410–1413 (2010).
- Testa, J. R. et al. Germline BAP1 mutations predispose to malignant mesothelioma. *Nat. Genet.* **43**, 1022–1026 (2011).
- Carbone, M. et al. BAP1 and cancer. *Nat. Rev. Cancer* **13**, 153–159 (2013).
- Murali, R., Wiesner, T. & Scolyer, R. A. Tumours associated with BAP1 mutations. *Pathology* **45**, 116–126 (2013).
- Jiao, Y. et al. Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. *Nat. Genet.* **45**, 1470–1473 (2013).
- LaFave, L. M. et al. Loss of BAP1 function leads to EZH2-dependent transformation. *Nat. Med.* **21**, 1344–1349 (2015).
- Comet, I., Riising, E. M., Leblanc, B. & Helin, K. Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat. Rev. Cancer*, **16**, 803–810 (2016).
- Govaere, O. et al. Keratin 19: a key role player in the invasion of human hepatocellular carcinomas. *Gut* **63**, 674–685 (2014).
- Yamashita, T. et al. EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features. *Gastroenterology* **136**, 1012–1024 (2009).
- Woo, H. G. et al. Identification of a cholangiocarcinoma-like gene expression trait in hepatocellular carcinoma. *Cancer Res.* **70**, 3034–3041 (2010).
- Ma, S. et al. Identification and characterization of tumorigenic liver cancer stem/progenitor cells. *Gastroenterology* **132**, 2542–2556 (2007).
- Matatall, K. A. et al. BAP1 deficiency causes loss of melanocytic cell identity in uveal melanoma. *BMC Cancer* **13**, 371 (2013).
- Shi, D. & Grossman, S. R. Ubiquitin becomes ubiquitous in cancer: emerging roles of ubiquitin ligases and deubiquitinases in tumorigenesis and as therapeutic targets. *Cancer Biol. Ther.* **10**, 737–747 (2010).
- Wu, X. et al. Ubiquitin-conjugating enzyme Ubc13 controls breast cancer metastasis through a TAK1-p38 MAP kinase cascade. *Proc. Natl Acad. Sci. USA* **111**, 13870–13875 (2014).
- Liu, J. et al. Targeting the ubiquitin pathway for cancer treatment. *Biochim. Biophys. Acta* **1855**, 50–60 (2015).
- Ben-Porath, I. et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.* **40**, 499–507 (2008).
- Budhu, A. et al. Prediction of venous metastases, recurrence, and prognosis in hepatocellular carcinoma based on a unique immune response signature of the liver microenvironment. *Cancer Cell* **10**, 99–111 (2006).
- Villa, E. et al. Neoangiogenesis-related genes are hallmarks of fast-growing hepatocellular carcinomas and worst survival. Results from a prospective study. *Gut* **65**, 861–869 (2016).
- Hoshida, Y. et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N. Engl. J. Med.* **359**, 1995–2004 (2008).

40. Yamashita, T. et al. EpCAM and  $\alpha$ -fetoprotein expression defines novel prognostic subtypes of hepatocellular carcinoma. *Cancer Res.* **68**, 1451–1461 (2008).
41. Kim, S. M. et al. Sixty-five gene-based risk score classifier predicts overall survival in hepatocellular carcinoma. *Hepatology* **55**, 1443–1452 (2012).
42. Kim, J. H. et al. Genomic predictors for recurrence patterns of hepatocellular carcinoma: model derivation and validation. *PLoS Med.* **11**, e1001770 (2014).
43. Roessler, S. et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res.* **70**, 10202–10212 (2010).

### Acknowledgements

This research was supported by grants from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (HI14C3395, HI5C1551, and HI16C1074), and the National Research Foundation of Korea (NRF) funded by the Korea government (MSIP) (NRF-2012R1A5A2048183).

### Author contributions

H.G.W. performed analysis and wrote the manuscript, which was edited by all co-authors. J.-H.C. and B.A.J. performed analysis. S.Y. performed experiments. E.J.C., J.-H.L. S.J.Y., J.-H.Y. N.-J.Y., K.-W.L., and K.-S.S. provided tissues and performed analysis. Y.J.K. contributed to overall study design and directed the study.

### Additional information

**Supplementary Information** accompanies this paper at doi:10.1038/s41467-017-00991-w.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017