Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma

Kan, Zhengyan; Zheng, Hancheng; Liu, Xiao; Li, Shuyu; Barber, Thomas D.; Gong, Zhuolin; Gao, Huan; Hao, Ke; Willard, Melinda D.; Xu, Jiangchun; Hauptschein, Robert; Rejto, Paul A.; Fernandez, Julio; Wang, Guan; Zhang, Qinghui; Wang, Bo; Chen, Ronghua; Wang, Jian; Lee, Nikki P.; Zhou, Wei; Lin, Zhao; Peng, Zhiyu; Yi, Kang; Chen, Shengpei; Li, Lin; Fan, Xiaomei; Yang, Jie; Ye, Rui; Ju, Jia; Wang, Kai; Estrella, Heather; Deng, Shibing; Wei, Ping; Qiu, Ming; Wulur, Isabella H.; Liu, Jiangang; Ehsani, Mariam E.; Zhang, Chunsheng; Loboda, Andrey; Sung, Wing Kin; Aggarwal, Amit; Poon, Ronnie T.; Fan, Sheung Tat; Wang, Jun; Hardwick, James; Reinhard, Christoph; Dai, Hongyue; Li, Yingrui; Luk, John M.; Mao, Mao

Published in:
Genome Research

DOI:
10.1101/gr.154492.113

Publication date:
2013

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma
Zhengyan Kan, Hancheng Zheng, Xiao Liu, et al.

Genome Res. 2013 23: 1422-1433 originally published online June 20, 2013
Access the most recent version at doi:10.1101/gr.154492.113

Supplemental Material
http://genome.cshlp.org/content/suppl/2013/07/15/gr.154492.113.DC1.html

References
This article cites 65 articles, 17 of which can be accessed free at:
http://genome.cshlp.org/content/23/9/1422.full.html#ref-list-1

Open Access
Freely available online through the Genome Research Open Access option.

Creative Commons License
This article, published in Genome Research, is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at http://creativecommons.org/licenses/by-nc/3.0/.

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to Genome Research go to:
http://genome.cshlp.org/subscriptions

© 2013, Published by Cold Spring Harbor Laboratory Press
Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma

Zhengyan Kan,1,13 Hancheng Zheng,2,13 Xiao Liu,2,3,13 Shuyu Li,4,13 Thomas D. Barber,4 Zhuolin Gong,2 Huan Gao,2 Ke Hao,5 Melinda D. Willard,4 Jiangchun Xu,1 Robert Hauritschein,1 Paul A. Rejto,1 Julio Fernandez,1 Guan Wang,2 Qinghui Zhang,2 Bo Wang,2 Ronghua Chen,5 Jian Wang,4 Nikki P. Lee,6 Wei Zhou,5 Zhao Lin,2 Zhiyu Peng,2 Kang Yi,2 Shengpei Chen,2 Lin Li,2 Xiaomei Fan,2 Jie Yang,2 Rui Ye,2 Jia Ju,2 Kai Wang,1 Heather Estrella,1 Shibing Deng,1 Ping Wei,1 Ming Qiu,1 Isabella H. Wulur,4 Jiangang Liu,4 Mariam E. Ehsani,4 Chunsheng Zhang,5 Andrey Loboda,5 Wing Kin Sung,6,7 Amit Aggarwal,4 Ronnie T. Poon,6 Sheung Tat Fan,6 Jun Wang,2,3,8,9 James Hardwick,5,10 Christoph Reinhard,4 Hongyue Dai,5 Yingrui Li,2,14 John M. Luk,6,11,12,14 and Mao Mao1,10,14

1Pfizer Oncology, San Diego, California 92121, USA; 2BGI-Shenzhen, Shenzhen 518083, China; 3Department of Biology, University of Copenhagen, Copenhagen, Denmark; 4Eli Lilly and Company, Indianapolis, Indiana 46285, USA; 5Merck Research Laboratories, Boston, Massachusetts 02115, USA; 6Department of Surgery, University of Hong Kong, Hong Kong, China; 7School of Computing, National University of Singapore, Singapore 117417; 8The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark; 9King Abdulaziz University, Jeddah, Saudi Arabia; 10Asian Cancer Research Group, Inc., Wilmington, Delaware 19808, USA; 11Departments of Pharmacology and Surgery, National University of Singapore, Singapore 117597; 12Institute of Molecular and Cell Biology, A*STAR, Singapore 138673

Hepatocellular carcinoma (HCC) is one of the most deadly cancers worldwide and has no effective treatment, yet the molecular basis of hepatocarcinogenesis remains largely unknown. Here we report findings from a whole-genome sequencing (WGS) study of 88 matched HCC tumor/normal pairs, 81 of which are Hepatitis B virus (HBV) positive, seeking to identify genetically altered genes and pathways implicated in HBV-associated HCC. We find beta-catenin to be the most frequently mutated oncogene (15.9%) and \textit{TP53} the most frequently mutated tumor suppressor (35.2%). The Wnt/beta-catenin and JAK/STAT pathways, altered in 62.5% and 45.5% of cases, respectively, are likely to act as two major oncogenic drivers in HCC. This study also identifies several prevalent and potentially actionable mutations, including activating mutations of Janus kinase 1 (\textit{JAK1}), in 9.1% of patients and provides a path toward therapeutic intervention of the disease.

[Supplemental material is available for this article.]
vating mutations in Janus kinase 1 (JAK1) that could lead to an opportunity to evaluate existing JAK inhibitors for the treatment of this deadly disease.

Results

Overview of genetic alterations

Eighty-eight primary HCC tumors and matched adjacent non-tumor liver tissues were analyzed by whole-genome DNA sequencing to identify somatic mutations and HBV integration sites. The vast majority (92%, n = 81) of patients in this cohort were HBV carriers (i.e., HBsAg seropositive) suffering from chronic hepatitis B or cirrhosis. None of the patients were hepatitis C virus (HCV) positive (see Supplemental Table 1).

For each DNA sample except for 11 patients where either normal or tumor DNA is limited, we built two libraries with different insert sizes (170 bp and 800 bp) for sequencing because the larger insert library provides better physical coverage for capturing SVs. The average depth of base pair coverage was 36.1× except for three tumor/normal pairs sequenced at 100× coverage. The concordance of single nucleotide polymorphism (SNP) calls between WGS and genotype data generated from Illumina 550K arrays exceeded 98.8%.

We developed a method to detect somatic SNVs in WGS data (see Methods) and estimated its sensitivity to be 82.2% (88/107) when evaluated against a reference set of 107 somatic SNVs identified from whole-transcriptome sequencing and experimentally validated using Sequenom mass spectrometric genotyping (see Methods). To assess the false-positive rate, we randomly selected 497 somatic SNV calls, and 3.4% were invalidated by Sequenom method in both tumor and normal DNA samples (Supplemental Table 2). Moreover, our validation results showed that sensitivity was not much higher in 100× samples (87.5%, 42/48) than in 30× samples (78%, 46/59) while false-positive rates were similar in the two groups (2.2% vs. 4.1%). Hence, we selected 30× as sequencing coverage for the study because it is more cost-effective for surveying the landscape of somatic SNVs in a large cohort than 100×. In total, 823,835 somatic SNV mutations were detected in 88 tumors, of which 5015 (0.6%) are predicted to affect protein coding (4482 was not much higher in 100×). The nonsynonymous to synonymous somatic SNV ratio of 2.99 and the mean protein-altering mutation rate of 1.8 per Mb are midrange among different cancer types (Greenman et al. 2007). The mutation prevalence and patterns observed are largely consistent with reports from a recent WGS study of a single, HCV-associated HCC genome (Supplemental Fig. 1; Totoki et al. 2011).

Using a somatic indel calling pipeline based on the SIRI method (Zhang et al. 2011) (see Methods), we also predicted 35,168 small (≤20-bp) deletions and 22,826 small insertions including 302 events affecting protein-coding (256 frame-shifted, 46 in-frame) in 285 genes (Supplemental Table 5). Eighty-eight of the 90 randomly selected 1- to 4-bp coding indels were validated (97.8%) by the Sequenom method in both tumor and normal DNA samples (Supplemental Table 6). In-frame deletions were found in oncogenes such as CTNNB1 (1.1%), MDM2 (1.1%), and IL6ST (1.1%) (Rebouissou et al. 2009), and frame-shifting deletions were found in tumor suppressors including ARID1A (2.3%), AXIN1 (2.3%), PTEN (1.1%), RB1 (1.1%), and TP53 (1.1%) (Supplemental Table 5).

We also detected 399 genomic HBV integration events affecting 115 coding genes and 4314 somatic structural variation (SV) events, including 260 gene fusions. Details of HBV integration and SV events have been described elsewhere (Sung et al. 2012; J Fernandez-Banet, NP Lee, KT Chan, H Gao, X Liu, WK Sung, W Tan, RT Poon, PA Reijto, M Mao, et al., in prep.).

Significantly mutated genes

The statistical significance of the observed mutation prevalence for each gene was assessed in the context of the background mutation rate and gene sequence length (Youn and Simon 2011) (see Methods and Supplemental Table 3). The most significantly mutated genes include three genes that have been previously established to be mutated in HCC (TP53, CTNNB1, AXIN1), two genes known to be frequently mutated in other cancer types (JAK1, LRPIB), and six genes not previously reported to be mutated in HCC (EPS15, SLC10A1, CACNA2D4, ALCY2, FAM5C, and COL11A1) (Table 1). TP53 has the highest prevalence of protein-altering mutations identified in our HCC cohort (31/88 samples = 35.2%), consistent with earlier HCC studies (Neuveut et al. 2010).

Table 1. Significantly mutated genes in primary HCC

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Mutation frequency</th>
<th>Confidence interval (95%)</th>
<th>No. COSMIC matched</th>
<th>No. recurrence matched</th>
<th>No. recurrence</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>Tumor protein p53</td>
<td>35.2% (31)</td>
<td>±10.0%</td>
<td>29</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>CTNNB1</td>
<td>Catenin (cadherin-associated protein), beta 1, 88 kDa</td>
<td>15.9% (14)</td>
<td>±7.6%</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>JAK1</td>
<td>Janus kinase 1</td>
<td>9.1% (8)</td>
<td>±6.0%</td>
<td>2</td>
<td>2</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>AXIN1</td>
<td>Axin 1</td>
<td>4.5% (4)</td>
<td>±4.4%</td>
<td>0</td>
<td>0</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td>SLC10A1</td>
<td>Solute carrier family 10 (sodium/bile acid cotransporter family), member 1</td>
<td>3.4% (3)</td>
<td>±3.6%</td>
<td>0</td>
<td>0</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td>CACNA2D4</td>
<td>Calcium channel, voltage-dependent, alpha2/delta subunit 4</td>
<td>5.7% (5)</td>
<td>±4.8%</td>
<td>0</td>
<td>0</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td>ADCY2</td>
<td>Adenylyl cyclase 2 (brain)</td>
<td>5.7% (5)</td>
<td>±4.8%</td>
<td>0</td>
<td>0</td>
<td>0.067</td>
<td></td>
</tr>
<tr>
<td>LRPIB</td>
<td>Low-density lipoprotein receptor-related protein 1B</td>
<td>11.4% (10)</td>
<td>±6.6%</td>
<td>0</td>
<td>0</td>
<td>0.073</td>
<td></td>
</tr>
<tr>
<td>FAM5C</td>
<td>Family with sequence similarity S, member C</td>
<td>5.7% (5)</td>
<td>±4.8%</td>
<td>0</td>
<td>0</td>
<td>0.077</td>
<td></td>
</tr>
<tr>
<td>COL11A1</td>
<td>Collagen, type XI, alpha 1</td>
<td>6.8% (6)</td>
<td>±5.3%</td>
<td>0</td>
<td>0</td>
<td>0.093</td>
<td></td>
</tr>
</tbody>
</table>

Shown above are significantly mutated genes with Benjamin-Hochberg false-discovery rate (FDR) ≤10%. The mutation frequency was defined to be the percentage of tumors with a protein-altering somatic mutation. Known cancer mutations were identified by comparing with somatic mutations in COSMIC database (Forbes et al. 2011), and the numbers of mutations with matching genomic coordinates were shown. Recurrent mutations are present in two or more HCC tumor samples.
We found TP53 mutated tumors harbor higher levels of chromosome instability, are more likely to be poorly differentiated, and have poor survival (Supplemental Fig. 2). This is consistent with the hypothesis that large-scale chromosomal alteration is left unchecked in cells where TP53-mediated defenses against DNA damage were deactivated by TP53 mutation (Chow and Poon 2010). Interestingly, the R249S mutation in TP53, a hotspot linked to aflatoxin B exposure in China (Aguilar et al. 1994), was observed in nine cases, suggesting that aflatoxin B exposure could be a secondary etiological agent in this cohort dominated by HBV+ cases and the mutation could be as the result of selective advantage during mutagenesis induced by aflatoxins and HBV chronic infection (Gous et al. 2009). CTNNB1 harbors protein-altering mutations in 15.9% of the tumors. Eleven of the 14 mutations are located in exon 3 and have been previously reported (Miyoshi et al. 1998). As exon 3 contains the phosphorylation site of GSK3B, a mediator of CTNNB1 degradation, prevalent mutations in this region may act to preserve CTNNB1 activities. We also detected a small deletion in exon 3 of CTNNB1 that removes the GSK3B phosphorylation sites (S33 and S37). In addition to CTNNB1, negative regulators of CTNNB1, including AXIN1 (4.5%), AXIN2 (2.3%), and APC (2.3%), are also mutated, implicating the canonical Wnt pathway as a major driver of hepatocarcinogenesis. Two known cancer genes not previously reported to be mutated in HCC are frequently mutated in our cohort—BRCAP (5.7%) and IGF1R (5.7%). Furthermore, there are multiple protein-altering mutations in cancer genes such as FLT3 (3.4%), KDR (3.4%), PTEN (2.3%), MET (2.3%), ALK (2.3%), ROS1 (2.3%), NOTCH2 (2.3%), and FGFR2 (2.3%). Mutations in PIK3CA and HRAS were observed but were infrequent (1/88). We did not observe mutations in KRAS, which has been reported in vinyl chloride-associated HCC (Wehrtrauch et al. 2001).

Several of the frequently mutated genes identified in this study are not well-established cancer genes, but corroborating evidence from recent studies points to potential roles in cancer. The chromatin remodeling gene ARID2 was found mutated in 14.0% of HCV-associated HCC, and 2.0% of HBV-associated HCC tumors recently (Li et al. 2011), and in our study, it was found mutated in three samples (3.4%) where all three mutations are nonsense. We also observed two mutations in ARID1A, a key component of the SWI-SNF chromatin remodeling complex reported to be frequently mutated in ovarian, colorectal, and gastric cancers (Wiegand et al. 2010; Jones et al. 2011; Wang et al. 2011). KMT2C (also known as MLL3), a member of the MLL gene family associated with aggressive acute leukemia (ALL) and a subunit of the histone H3K4 methyltransferase complex (Lee et al. 2009), is mutated in 8.0% of tumors. KMT2B, a paralog of KMT2C, is affected by HBV integration in 10.2% of tumors. Taken together, our findings suggest that chromatin remodeling may play a significant role in HCC carcinogenesis, echoing recent findings in medulloblastoma, renal carcinoma, gastric cancer, and ovarian cancer genome sequencing efforts (Wiegand et al. 2010; Varela et al. 2011).

In addition to well-known and emerging cancer genes, our study revealed several genes with poorly characterized functional roles in cancer but sufficient mutational evidence to warrant experimental investigation. One of the most significantly mutated genes is LRPIB with mutations in 10 tumors (11.4%) and LRPIB mutations have been previously reported in 9% of lung adenocarcinomas (Ding et al. 2008). EPS15, a phosphorylation substrate of EGFR and a translocation partner of MLL in acute leukemia, is significantly mutated with four mutations (4.5%). ADCY2 belongs to the adenylyl cyclase family of enzymes responsible for the synthesis of cAMP and is mutated in five cases (5.7%). USH2A mutations in eight tumors (9.1%) including two nonsense mutations were identified in this study, indicating that USH2A might function as a tumor suppressor. Germline mutations in the USH2A gene are responsible for the most common subtype of Usher syndrome (USH), a genetic disorder that is a main cause of combined deaf-blindness (Reiners et al. 2006). Interestingly, three additional USH-associated genes, GPR98 (8.0%), PCDH15 (6.8%), and MYO7A (1.1%), are also mutated in this cohort, putatively linking the genetic causes of Usher syndrome with hepatocarcinogenesis.

**JAK1 mutations and functional characterization**

JAK1 (Janus kinase 1) is a member of the Janus kinase family that mediates the growth factor and cytokine-induced STAT signaling pathway and plays important roles in immunity, cell growth, and differentiation. Activating mutations in JAK1 and JAK2 have been described in patients with various hematologic malignancies, including acute leukemia and myeloproliferative neoplasm (MPN) (Quintas-Cardama et al. 2011). We identified seven distinct protein-altering mutations in JAK1 in eight HCC tumors (9.1%), with two of the mutations (S703I and S729C) recurring in two tumors. All mutations were validated by Sanger sequencing. Six of the mutations are located in the pseudo-kinase (JH2) and tyrosine kinase (JH1) domains (Fig. 1A). The tandem mutation Q644H/V645F was recently reported in one HCC tumor (Xie et al. 2009), and S703I has been independently shown to activate JAK1 in a mouse model system characterizing IL-9/JAK signaling (Supplemental Fig. 3; Hornakova et al. 2011). We therefore examined whether the seven JAK1 somatic mutations identified in our study constitutively activate the JAK/STAT pathway. Upon expression of the mutants in HEK293FT and a HCC cell line Hep3B, we observed increased autophosphorylation of JAK1 (S703I and S729C) and increased activation of STAT3 (Q644H/V645F, S703I, S729C, and L910P) compared with wild-type JAK1 (Fig. 1B). No changes in signaling were observed with the FERM domain mutant N242S or with the kinase domain mutant G902E. Cytokines such as IL6 activate STAT signaling in HCC cell lines, including Hep3B, and the presence of such receptor complexes has been shown to mediate mutant JAK1 signaling (Hornakova et al. 2009; Pilati et al. 2011); thus, we sought to examine whether IL6 regulates variant JAK1 signaling in Hep3B cells. Although JAK1 Q644H/V645F, S703I, S729C, and L910P constitutively increase JAK/STAT signaling compared with wild-type, IL6-dependent regulation of the JAK1 mutants is not enhanced beyond wild-type activity (Fig. 1B). We modeled the interaction between the JH1 and JH2 domains using a crystal structure of the JH1 domain (3EYG) (Williams et al. 2009) and a homology model of the JH2 domain oriented to have an interaction between the first N-terminal α-helices in each domain, patterned after previous models (Fig. 1C; Flex et al. 2008). Notably, three of the mutated residues located in the pseudo-kinase domain that constitutively activate the JAK/STAT pathway (Q644H/V645F, S703I) are predicted by the model to be at the JH1/JH2 interface. Mutation of these residues could alleviate the inhibition of JAK1 kinase activity by the pseudo-kinase domain; a negative regulatory function for the JH2 domain has been established in Janus kinases (Saharinen and Silvernño 2002), and the A634D mutation in the JAK1 JH2 domain has been reported to up-regulate kinase activity ligand-independently (Flex et al. 2008).

To evaluate the endogenous functional consequences of JAK1 mutations, we performed gene expression profile analysis on the 88 HCC tumors using an independently derived gene signature.
from acute lymphoblastic leukemia patients with acquired JAK1 mutations that is enriched in genes known to be modulated by JAK/STAT signaling (Flex et al. 2008). All six HCC tumors harboring JAK1 activating mutations (Q644H/V645F, S703I, S729C, and L910P) exhibit a strong expression signature as quantified by principal component analysis (PCA), while the two tumors with nonactivating mutations (N242S and G902E) do not have the signature, suggesting a gain-of-function role for these activating JAK1 mutations in primary HCC tumors (Fig. 1D). To assess oncogenic consequences of the JAK1 activating mutations, we evaluated the ability of each to induce autonomous growth when expressed in cytokine-dependent Ba/F3 cells. Wild-type JAK1 or selected JAK1 mutants (S703I, S729C, and L910P) were transduced into Ba/F3 cells, and cell growth was assessed in the absence of the cytokine IL3. All three mutants conferred IL3–independent growth; in contrast, cells transduced with wild-type JAK1 remained dependent on the cytokine for survival, comparable to cells transduced with vector alone. Among the three activating mutations tested, it appeared that S729C exhibited the most transforming capability (Fig. 1E). JAK2 inhibitors have proven to be promising therapies for JAK2 mutation driven myeloproliferative neoplasms (Quintas-Cardama et al. 2011). We evaluated two mutant JAK1 transduced Ba/F3 cell lines for their sensitivity to the JAK1/2 dual inhibitor ruxolitinib (Quintas-Cardama et al. 2010), and the JAK2 inhibitor BMS-911453 (Purandare et al. 2012). Ruxolitinib displayed potent growth inhibition on both S703I and
S729C transduced cells, whereas BMS-911453 only showed a partial inhibitory effect (Fig. 1F), and growth inhibition by ruxolitinib correlated with its complete inhibition of downstream STAT3 phosphorylation (Supplemental Fig. 4). The differing cellular activity observed with the two JAK inhibitors may be due to ruxolitinib being a 100-fold more potent inhibitor of the JAK1 enzyme (i.e., JAK1 IC50 of 3.3 nM for ruxolitinib vs. 356 nM for BM-911453). Taken together, these results suggest that the JAK1 mutations identified in this study are potentially oncogenic drivers in HCC through constitutive activation of downstream signaling.

Copy number changes

The landscape of copy number variations (CNVs) in HCC genomes is predominated by large-scale amplifications and deletions of chromosomal arms or entire chromosomes (amplified: 1q, 5p, 6p, 8q, 17q, 20q, and Xq; deleted: 4p-q, 8p, 13p-q, 16p-q, 17p, 21p-q, and 22q) (Fig. 2A; Supplemental Table 7; Supplemental Fig. 5). We performed a GISTIC analysis adapted to WGS data and revealed 1967 genes in 104 amplified regions and 2604 genes in 262 deleted regions (Supplemental Tables 8, 9). In total, 7.8% of the genome is amplified with an average gain of 0.65 copies, and 10.5% of the genome is deleted with an average loss of 0.77 copies. In significantly amplified or deleted regions, the G-scores of amplification for oncogenes as classified by Cancer Gene Census (Futreal et al. 2004) are threefold higher on average than tumor suppressors (0.079/0.025), whereas G-scores of deletion for tumor suppressors are 2.7-fold higher than oncogenes (0.244/0.095) (Supplemental Fig. 6). Outside of CNV regions, however, there is little difference overall in G-scores between oncogenes and tumor suppressors. The selective amplification of oncogenes and deletion of tumor suppressor genes suggest that these broad copy number changes in HCC genomes likely contribute to hepatocarcinogenesis.

Applying a filter based on cis-correlation of DNA copy numbers and mRNA expression levels (see Methods), we focused on 685 candidate driver genes in amplified regions and 821 genes in deleted regions, including many known and putative oncogenes—MDM4 (1q), ELK4 (1q), PARP1 (1q), MYC (8q), PTK2 (8q), CCND1 (11q), FGFR9 (11q), RPTOR (17p), IRAK1 (Xq)—and tumor suppressors—ARID1A (1p), FBXW7 (4q), PTK2B (8p), CDKN2A/B (9p), RB1 (13q), AXIN1 (16p), TP53 (16q), and MAP2K4 (16q) (Fig. 2B). FGFR9 along with CCND1 were among the top two driver genes in HCC by a functional genomics screen (Sawey et al. 2011). We found that the CCND1/FGFR9 region is focally amplified in four samples (4.5%) with the highest amplitude among all genes, and the amplification is concordant with a significant increase in gene expression (Fig. 2C). RSP02, a member of the R-spondin family that activates beta-catenin signaling through LGR receptor (de Lau et al. 2011), had a similar pattern of overexpression driven by amplification. Conversely, well-known tumor suppressors such as TP53, RB1, CDKN2A, and CDKN2B reside in the most frequently deleted regions in HCC genomes. Other genes in regions that are frequently deleted include FBXW7, ARID1A, and MAP2K4, consistent with their putative roles as tumor suppressors.

Frequently altered cancer pathways

To further elucidate the functional impact of the complex alterations observed in HCC genomes, we performed an integrative analysis of somatic mutation, CNV, and HBV integration mapped to canonical cancer pathways from MSigDB (Liberzon et al. 2011), and identified frequent alterations across multiple pathways, including Wnt, JAK/STAT, G1/S cell cycle, and apoptosis (Supplemental Table 10; Supplemental Fig. 7). Pathway enrichment analysis (see Methods) further identified significant overlaps (FDR < 0.1) with cell adhesion, DNA repair, and chromatin modification pathways (Supplemental Table 10).

The Wnt pathway drives proliferation, maintains survival and promotes stemness in cancer cells (Klaus and Birchmeier 2008), and plays a well-known role in HCC tumorigenesis (Whittaker et al. 2010; Fatima et al. 2011). Overall, 62.5% of tumors in the cohort contain genomic alterations composed of SVN and CNV affecting core components of the Wnt pathway (Fig. 3A). Nonsynonymous somatic mutations occur in 23.9% of the tumors, including mutations in CTNNB1 (15.9%) and components of the destruction complex (APC and AXIN1/2) at a frequency of 8.0%. In addition, the tumor suppressor AXIN1 (Satoh et al. 2000) was deleted in 11.4% of tumors. Proximal to the destruction complex, amplifications were detected both in the pathway ligands (WNT9A, 25.0%; RSPO2, 21.6%) and in a pathway receptor (EFZD6, 21.6%). It remains to be determined which genomic lesions translate into aberrant Wnt pathway activation in these tumors as measured by nuclear beta-catenin staining or the presence of a Wnt pathway signature. While the majority of the tumors in this study harbor genetic lesions of Wnt pathway components, activation of Wnt signaling can occur in additional or overlapping tumor cases through down-regulation of SFRP1 and DAFT1 gene expression, known negative regulators of Wnt and Dvl family proteins, respectively (Kawano and Kypta 2003; Yau et al. 2005). Hepatocytes are capable of rapid regeneration and proliferation through activating the stem cell compartment in response to chronic liver damages inflicted by hepatitis, cirrhosis, and other hallmark conditions of HCC. An earlier study has shown that 28%–50% of HCCs express markers of liver progenitor cells, suggesting a link between HCC and these constantly proliferating stem cells (Mishra et al. 2009). Frequent mutations in the Wnt pathway revealed in this study, coupled with the observation that the recently identified R-spondin receptor LGR5 (de Lau et al. 2011), an adult stem cell marker and beta-catenin target, is up-regulated in HCC further strengthens the hypothesis that HCC may have a significant stem cell lineage.

The pathogenesis of HCC is frequently linked to inflammatory responses triggered by chronic viral infection and cell necrosis. Cytokine-induced JAK/STAT pathway activation can lead to tumor-promoting inflammation in HCC, where the majority of cases were shown to harbor activated STAT3 (He and Karin 2011). Our findings of prevalent and activating JAK1 mutations shed light on a heretofore undervalued role for JAK1 in promoting STAT3 signaling and hepatocarcinogenesis. Further, as much as 45.5% of the HCC tumors harbor genomic alterations in the JAK/STAT pathway, including frequent amplification and mutation of the cytokines PRL, IL7, and IL20, and the cytokine receptors IL6ST, IL6R, and IL21R (Fig. 3B), although it remains to be determined if these genetic alterations would lead to JAK/STAT pathway activation. It has been demonstrated experimentally that GNAS mutations activate STAT3 to induce inflammatory responses in a rare subtype of inflammatory liver cancer, suggesting cross-talk of the cAMP and JAK/STAT pathways (Nault et al. 2011). We noted that adenylate cyclase type 2 (ADCY2), an important member of the cAMP synthesis pathway activated by GNAs, is mutated in 5.7% of this HCC cohort. Furthermore, GNAS and four other members of the adenylate cyclase family (ADCY3, -8, -9, and -10) are also mutated in one case each. Taken together, our findings suggest that the JAK/STAT pathway, potentially acting in concert with the
Figure 2. Landscape of genomic copy number alteration in HCC. (A) Shown is a genome-scale overview of copy number segments across 88 HCC samples rendered by IGV (Robinson et al. 2011). Each track in the upper panel represents one tumor sample, and the lower panels show the amplification and deletion G-score profiles across the genome. (B) Shown is a chromosomal overview of amplification and deletion peaks with the locations of putative cancer driver genes indicated. Arrows indicate focal peaks (<100 kb). Asterisks indicate genes also harboring nonsense mutations. (C) A log-scale plot of gene-level expression fold-change in CNV relative to non-CNV samples versus amplitude of copy gains. Genes with strong increase in expression driven by copy gains are labeled. Colors and shadings were used to indicate the CNV and cis-correlation statuses of genes. The size of the marker is proportional to the amplification G-score.
G-protein/cAMP pathway, may induce inflammatory responses that promote carcinogenesis in a substantial subset of HCC.

HCC genomes suffer extensive damage in the form of large-scale copy number alterations and viral integrations, which if left unchecked, would be expected to trigger TP53-mediated apoptosis and cell cycle arrests. Frequent mutations and deletions of TP53 appear to have disabled this important line of cellular defense in an aggregate of 43.2% of HCC tumors, which are more chromosomal instable and have poor survival. We found that up to 42.0% of HCC tumors harbor genomic alteration in the genes operating the G1/S check point of the cell cycle (Fig. 3C). RB1, a key inhibitor of cell cycle progression, harbors nonsense mutations in three cases (3.4%) and genomic deletion in 11 cases (12.5%). CDKN2A and CDKN2B, which negatively regulate CDK activities, are deleted in 10.2% and 12.5% of the cases, respectively. On the other hand, promoters of cell cycle progression are activated by DNA amplification (CCND1, 4.5%) and HBV integration (CCNE1, 4.5%), resulting in markedly higher gene expression. The apoptosis pathway, an executor of cell death, could be deactivated by DNA copy number changes in as much as 45.5% of HCC tumors (Fig. 3D). Death receptors such as TNFRSF10A/B are deleted in 22.7% of cases, and downstream signaling is further disrupted by genomic deletions of TRADD, CASP3, CASP9, DFFA, and DFFB. In contrast, XIAP, an inhibitor of caspases and apoptosis is amplified in 8.0% of tumors. Hence, the functions of TP53-related pathways such as the G1/S check point and apoptosis also appear to be genetically altered through selective activation and deactivation of key regulators.

HCC molecular classification

Three HCC subclasses (S1, S2, and S3) have been identified based on a meta-analysis of nine gene expression profiling studies (Hoshida et al. 2009). While gene expression signatures and some clinical phenotypes for each subclass have been described, the underlying genetic alterations remain largely unknown. We also clustered 88 HCC tumors in our study into three subclasses based on gene expression array data and applying supervised methods using previously defined gene signatures (Fig. 4A; Hoshida et al. 2009). To delineate the genetic basis of the HCC subclasses, SNV, CNV, and HBV integration data for significantly altered genes and pathways were mapped to three subclasses, revealing distinct genetic profiles for each subclass (Fig. 4B,D; Supplemental Fig. 8, Supplemental Tables 12, 13). S1 and S2 express high level of genes involved in cell cycle control and cell proliferation. Most S1 and S2 tumors are poorly or moderately differentiated with a high rate of recurrence. A subset of S1 and S2 tumors harbors HBV integration into the KMT2B gene locus. S1 tumors also express a high level of...
genes in immune response and angiogenesis. S2 has the highest frequency of \textit{TP53} mutation and the highest serum AFP level. S3 tumors are well or moderately differentiated with a gene expression profile reflecting normal liver function. In addition, S3 has relatively high frequency of \textit{CTNNB1} and \textit{JAK1} mutations and has HBV integration into the \textit{TERT} gene locus. Integrative analysis of gene expression profiles, genetic alterations, and clinical characteristics appears to at least partially explain the observed difference in progression-free survival of these subclasses (Fig. 4C).

**Potentially actionable genetic alterations**

Our findings of several mutations and gene amplifications in HCC have clinical implications (Table 2). Inhibitors of these enzymes or key components of the pathways are already in clinical development for non-HCC indications. The prevalence of these genetic alterations in HCC discovered from our study warrants testing some of these existing inhibitors in HCC preclinical models and patients. Recurrent \textit{JAK1} mutations in HCC have not been reported previously. Nine percent of HCC tumors in this cohort have \textit{JAK1} mutations. The majority of mutations reside in the pseudo-kinase and kinase domains and are functionally activating, and mutant-induced cell growth can be suppressed by the JAK1/2 dual inhibitor ruxolitinib. Our data suggest ruxolitinib, recently approved by the FDA to treat myelofibrosis (Mesa et al. 2012), should be considered for testing in HCC patients with activating \textit{JAK1} mutations.

Preclinical studies have shown that cell lines with elevated cyclin D1 and decreased CDKN2A (also known as p16) expression are most

![Figure 4](https://example.com/figure4.png)

**Figure 4.** Molecular subclassification of HCC. (A) Gene expression profile of previously defined signatures (Hoshida et al. 2009) in three HCC subclasses. High and low expressions are represented by red and blue color, respectively. Seventy-six of the 88 tumors (86%) are assigned to subclasses with high confidence (FDR < 0.05). (B) Genetic and clinical profiles of the three HCC subclasses. Tumors with somatic SNVs in \textit{CTNNB1}, \textit{TP53}, and \textit{JAK1}, HBV integrations into \textit{TERT} and \textit{MLL4} (also known as \textit{KMT2B}), AFP level >400 ng/mL, or tumor recurrence are represented by dark blue color. For tumor grade, poor, moderate, and well differentiation are represented by dark blue, light blue, and gray color, respectively. Tumors with missing tumor grade information are shown in white. (C) Kaplan-Meier survival plot for the three HCC subclasses. (D) Schematic summary of gene expression, genetic, and clinical profiles for each HCC subclass.

### Table 2. Potentially actionable mutations and matched clinical stage inhibitors

<table>
<thead>
<tr>
<th>Genes</th>
<th>Mutation</th>
<th>Amplification</th>
<th>Deletion</th>
<th>Combined frequency</th>
<th>Inhibitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{JAK1}</td>
<td>\textit{JAK1}</td>
<td>—</td>
<td>—</td>
<td>9.1%</td>
<td>\textit{JAKi} (ruxolitinib)</td>
</tr>
<tr>
<td>\textit{FAK}</td>
<td>—</td>
<td>\textit{FAK}</td>
<td>—</td>
<td>26.1%</td>
<td>\textit{FAKi} (PF-045,54878, PF-562271)</td>
</tr>
<tr>
<td>\textit{CCND1}, \textit{CDKN2A}</td>
<td>—</td>
<td>\textit{CCND1}</td>
<td>\textit{CDKN2A}</td>
<td>4.5%</td>
<td>14.7%</td>
</tr>
<tr>
<td>\textit{FGF19}</td>
<td>—</td>
<td>\textit{FGF19}</td>
<td>—</td>
<td>4.5%</td>
<td>\textit{FGFri} (brivanib, BQ398, LY2874455)</td>
</tr>
<tr>
<td>\textit{BRCA1/2}, \textit{PARP1}</td>
<td>\textit{BRCA1} (1.1%), \textit{BRCA2} (5.7%), \textit{PARP1} (18.2%)</td>
<td>—</td>
<td>—</td>
<td>25.0%</td>
<td>\textit{PARPi} (AG-14699, olaparib)</td>
</tr>
</tbody>
</table>
sensitive to cyclin D kinase 4/6 inhibitors (Linn et al. 2009). PD-0332991, a selective CDK4/6 inhibitor, is being tested in advanced breast cancer with CCND1 amplification and/or loss of CDKN2A (p16). We observed 4.5% HCC tumors with CCND1 focal amplification and 10.2% with CDKN2A deletion. Our data suggest this subset of HCC tumors could be candidates of CDK4/6 inhibitor trials.

Two focal adhesion kinase (FAK, also known as PTK2) inhibitors are in phase 1 clinical trials and have demonstrated modest clinical benefits in a small unselected population. FAK amplification has been associated with poor survival in breast and gastric cancers (Park et al. 2011; Yon et al. 2011). With 26.1% of HCC tumors harboring FAK amplifications, our data suggest testing FAK inhibitors in the FAK amplified HCC models.

PARP inhibitors such as olaparib have been showing efficacy in BRCA-deficient ovarian and breast cancer trials (Audeh et al. 2010; Tutt et al. 2010). We find the DNA repair pathway extensively altered in HCC with PARP1 amplified in 18.2% of cases and mutations in BRCA1 (1.1%) and BRCA2 (5.7%). Though these alterations remain to be functionally tested, our findings provide empirical evidence for the inclusion of HCC in the PARP inhibitor trials.

The fibroblast growth factor ligand (FGF19) is amplified in 4.5% of HCC tumors. A neutralizing anti-FGF19 antibody blocks clonogenicity and tumorigenicity of HCC models harboring FGF19 amplification (Sawey et al. 2011). As a proof-of-concept study, we tested the FGFR small-molecule inhibitor PD173074 in a panel of 14 HCC cell lines for inhibition of cell proliferation. Indeed those HCC cell lines harboring FGFR4/FGF19 amplification and overexpressing FGFR3/FGFR4 display increased sensitivity to PD173074 (Supplemental Table 14). Based on our data and the emerging preclinical data, we propose to test FGFR inhibitors including brivanib, a dual VEGFR/FGFR inhibitor, which is being evaluated in phase 3 HCC trials, in HCC patients with FGF amplifications and FGFR overexpressions.

Discussion

The results of this study highlight important differences between HCC and other solid tumors such as lung, breast, prostate, and colon cancers. While EGFR, PI3K, and MAPK pathway alterations are common in other cancer types, Wnt/beta-catenin and JAK/STAT are the two major oncogenic pathways in HCC. These findings may explain why some targeted therapies that are effective in other tumor types have not demonstrated significant improvement on HCC patient survival. Cataloging major mutations and pathways by cancer genome sequencing can lead to a better understanding of tumor biology and the discovery and prioritization of drug targets. The time line of drug development may be shortened when compounds already in late-stage preclinical or clinical development for other indications are available for testing in HCC (Chin et al. 2011). Activating JAK1 mutations discovered from this study fit into the class of so-called actionable mutations, as do amplifications of FAK, CCND1, FGF19, and PARP, and inactivating mutations in BRCA1 and 2. Most importantly, this study demonstrates that moribundicentric genome sequencing is not only a powerful tool to study disease biology and identify new drug targets, but also provides opportunities for promptly translating scientific discoveries into improvements in patient care.

Methods

Samples and genomic analyses

Sample collection, gene expression array analysis, SNP array analysis, and whole-genome sequencing have been described in previous publications (Lamb et al. 2011; Sung et al. 2012). Microarray data have been deposited in the GEO database under accession numbers GSE28127 (for SNP arrays) and GSE25097 (for expression arrays). Sequence data have been deposited in the European Nucleotide Archive (ENA, http://www.ebi.ac.uk/ena/) under accession number ERP001196 and in GigaDB (http://dx.doi.org/10.5524/100034).

Somatic mutation detection

Sequencing reads were aligned to the hg19 reference genome using SOAP2 (Li et al. 2009) followed by removal of PCR duplicates, low-quality (Q < 20), and nonuniquely mapped reads. Mutations in tumor were first predicted by SOAPsnp (http://soap.genomics.org.cn/SOAPsnp.html) using a sensitive score threshold, and a P-value was calculated using Fisher’s exact test for all putative mutation sites based on the distribution of read support for different alleles in tumor and matched normal samples (Supplemental Fig. 9). A somatic mutation is called if the following criteria were met: (1) read depth ≥10 in both tumor and normal samples; (2) read support of mutant allele in tumor tissue not a result of sequencing error (binomial test, f = 0.1, P > 0.01); (3) quality score not significantly lower than other alleles (Wilcoxon rank sum test, P > 0.01); (4) mutant allele frequency change between tumor and adjacent normal ≥20% and Fisher’s exact test P < 0.01; (5) mutant allele not significantly enriched in repeatedly aligned reads; and (6) mutant allele not significantly enriched within 10 bp of 5’ or 3’ ends of reads (Fisher’s exact test, P > 0.01).

Indels in tumor samples were predicted by the split reads tool (SRiC) method (Zhang et al. 2011). Gapped alignment was performed using BWA aligner (Li and Durbin 2009), allowing only one gap located >5 bp to either read ends for each mapped read. To identify somatic indels, all reads from normal samples were aligned to sequence templates representing reference and variation genotypes for each detected tumor indel. Tumor indels with six or more matched normal reads as well as those matching dbsNP indel records were filtered as germline indels.

Somatic SNV and indel validation

Genotyping assays were performed via iPLEX MassARRAY system (Sequenom). Both PCR and MassEXTEND primers for each mutation were in silico designed by MassARRAY Assay design 4.0 software. Multiplex PCR were carried out by GeneAmp PCR System 9700 Dual 384-Well Sample Block Module (Applied Biosystems), followed by dephosphorylation, single base extension reaction, and desalting. MassARRAY Nanodispenser RS1000 was used to automatically spot reactions to 384 SpectroCHIP, which was further placed into the MALDI-TOF mass spectrometer. All genotype calls by MassARRAY Typer 4.0 were manually confirmed by examining the spectra for each assay and sample.

Significantly mutated genes analysis

We used the method described previously by Youn and Simon (2011) to compute the significance of observed mutations on each gene. The statistical model takes both mutation prevalence and functional impact into consideration. Functional impact was evaluated as mutation score assigned in the following order: missense < inframe indel < mutation in splice sites < frameshift indel = nonsense. Different types of missense mutations were also assigned different scores based on BLOSUM80 matrix. This method assumes that passenger mutations, including silent and nonsilent mutations, were generated from the same background mutation process. Incorporating different background mutation rates of each sample,
Hepatocellular carcinoma whole-genome sequencing

background distribution of mutation score for each gene was computed. The P-value for each gene was calculated from this background distribution and the test statistics from the observed mutation scores across samples. The P-value was adjusted using the Benjamini-Hochberg method to estimate the false-discovery rate (FDR). Significantly mutated genes were selected if FDR ≤ 10%.

Validation of JAKI mutations

We designed 19 pairs of primers that uniquely amplify 24 exons of JAK1 gene using Primer Premier 5.00 (Premier Biosoft) and Oligo 6.71 (Molecular Biology Insights). PCR was performed on Dual 96-well GeneAmp PCR System 9700 (Applied Biosystems). The 50-μL reaction contains 1× PCR buffer, 0.2 mM dNTP mixture, 200 nM forward and reverse primers, 2.5 units Taq (Takara Bio), and 20 ng template DNA. Cycling conditions were 1 min at 94°C for initial denaturation of the DNA and polymerase activation, followed by 35 cycles of 30 sec denaturation at 95°C, 30 sec annealing at 62°C–67°C, and 30 sec – 90 sec extension at 72°C. The products were sequenced by 3730xl DNA Analyzer (Applied Biosystems). All sequences were analyzed by the Sequencing Analysis Software version 5.2 (Applied Biosystems).

Experimental characterization of JAKI mutations

Flag-tagged human JAK1 ORFs for wild-type and mutant variants (N242S, Q644H/V645F, S703I, S729C, G902E, and L910P) were cloned into pcDNA3.1(+)(Invitrogen). Anti-M2 Flag HRP and -beta actin (Sigma), anti-pJAK1 (pY1034/1035), -pSTAT3 (pT705), and -STAT3 (Cell Signaling) were used for Western blotting. All cell culture reagents were from Hyclone. All cells were cultured at 37°C in 5% CO2 humidified air. HEK293FT cells were maintained in DMEM (high glucose) supplemented with 10% fetal bovine serum, 4 mM L-glutamine, 1 mM sodium pyruvate. Hep3B cells were maintained in MEM/EBSS supplemented with 10% fetal bovine serum, 0.1 mM nonessential amino acids, 1 mM sodium pyruvate. For Western blotting experiments, HEK293FT and Hep3B cell cultures were seeded the day before at 800,000 and 400,000 cells/well, respectively, in a six-well dish and transfected with a total of 1.5 μg plasmid DNA using Fugene 6 (Roche) or FuGENE HD (Promega) at a ratio of 3:1 (Fugene:DNA).

HEK293FT and Hep3B cells were starved for 48 h post-transfection in 0% serum for 4 h. Cells were lysed on ice in cold RIPA buffer (150 mM NaCl, 1% Igepal CA-630, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris at pH 8.0; Sigma) plus protease and phosphatase inhibitors (HALT protease and phosphatase inhibitor cocktail; Pierce). Cell lysates were sonicated in an ice water bath for 5 min and then centrifuged at 16,500 × g for 20 min at 4°C. A portion of the supernatant was removed, mixed 1:1 with 6× Laemmli sample buffer, and boiled for 5 min. Samples were subjected to SDS-PAGE and transferred to nitrocellulose using the iBlot dry blotting system (Invitrogen). Membranes were blocked in 5% nonfat milk/TBS-T for 1 h at room temperature. Western blotting was performed using the aforementioned primary antibodies in 3% BSA/TBS-T, secondary anti-mouse or -rabbit IgG antibody-HRP conjugates in 1% BSA/TBS-T (GE Healthcare), and enhanced chemiluminescence (SuperSignal West Pico or SuperSignal West Femto Pierce).

Human JAK1 ORFs for wild-type and mutant variants (N242S, S703I, S729C, G902E, and L910P) were constructed in the bicistronic retroviral vector pMX-IRES-GFP (Cell Biologs). Retroviral supernatants were used to infect Ba/F3 cells cultured in RPMI 1640 medium containing 10% FBS and 10% WEHI-3B cell condition media. GFP-positive cells were isolated by flow-cytometric sorting and subsequently expanded. Equal GFP expression levels of transduced cells were confirmed by FACS analysis. For assaying cytokine independence, transduced Ba/F3 cells were cultured in the absence of IL3. Cell viability was determined by Trypan blue exclusion.

For compound treatment, mutant JAK1 transduced Ba/F3 cells were plated on 96-well plates (2000 cells/well) in RPMI 1640 medium containing 2% FBS. They were treated with compounds at indicated concentrations for 3 d. Cell viability was measured using CellTiter-Glo Luminescent Cell Viability Assay (Promega) following the manufacturer’s instructions.

For Western blot analysis, JAK1 (S703I)-transduced Ba/F3 cells were treated with 1 μM compounds for 2 h in serum-free media. Approximately 30 μg of lysate were resolved by electrophoresis through a polyacrylamide gel (BioRad Laboratories). Following transfer to nitrocellulose membrane, the membrane was subjected to immunoblot analysis using pSTAT3 (pY705) and STAT3 antibodies (Cell Signaling), and mouse α-actin antibody (Sigma-Aldrich). Secondary antibodies were purchased from Licor Biosciences. The images were analyzed using Odyssey Imager (Licer Biosciences).

JAKI pseudo-kinase and tyrosine kinase domain interaction modeling

We modeled the interaction between the pseudo-kinase and tyrosine kinase domains using a crystal structure of the tyrosine kinase domain (3EYG) (Williams et al. 2009) and a homology model of the pseudo-kinase domain oriented to have an interaction between the first N-terminal α-helices in each domain, patterned after previous models (Flex et al. 2008) using Discovery Studio (accelrys.com).

JAKI gene signature analysis

The JAK1 mutation gene signature was derived from acute lymphoblastic leukemia patients (Flex et al. 2008). The 112 differentially expressed genes were mapped to expression profiles of our cohort. Using principal component analysis (PCA), we then computed the first principal component value for each sample as the JAKI activation score.

CNV analysis

We used SegSeq (Chiang et al. 2009) to infer somatic copy number variation (CNV) in HCC genomes using WGS reads. The resulting copy number segments were mapped to individual genes to determine gene-level copy numbers and copy gain/loss statuses using thresholds of ≥3 copies for gain and ≤1.25 for loss. To infer recurrently amplified or deleted genomic regions, we reimplemented the GISTIC algorithm (Bereukhim et al. 2007) using copy numbers in 1-kb windows instead of SNP array probes as markers. G-scores were calculated for genomic and gene-coding regions based on the frequency and amplitude of amplification and deletion affecting the gene respectively. A significant CNV region is defined as having a G-score > 1.25 for amplified and > 0.8 for deleted genes.

Canonical pathway analysis

Thirty-three canonical cancer pathways were selected from MSigDB (Liberzon et al. 2011). For each gene and tumor sample,
also called an event, a binary status (1, altered; 0, no change) is determined for somatic mutation, cis-correlated CNV, and HBV integration. The integrated alteration status is simply a sum of three alteration statuses. Pathway alteration prevalence is represented at the sample-level as the proportion of tumors harboring alteration in any pathway gene. In addition, prevalence is calculated at the event level as the percentage of events altered and gene level as the percentage of pathway genes harboring one or more alterations.

Pathway enrichment analysis

Pathway enrichment analyses of genes harboring somatic SNV or CNV were performed using the KEGG or Gene Ontology canonical pathways. P-values were calculated based on hypergeometric distribution with FDR correction using the Benjamini and Hochberg method (Hochberg and Benjamini 1990). GO term enrichment analysis was performed using only the FAT terms and DAVID (Dennis et al. 2003).

HCC molecular classification

A previously defined 619-gene signature for three HCC subclasses and binary weights for each gene were used to define cluster centroids. Each of the 88 HCC samples in the current study was assigned to the closest centroid based on expression of the 619-gene signature. A permutation test was used to assess the statistical significance, and an FDR of 0.05 was used to define high-confidence class assignment. To identify genes overexpressed in each subclass, t-tests were carried out to compare gene expressions in each subclass with those in the rest of the HCC samples. A P-value <0.05 and fold change >1.5 were used to define statistical significance. Significantly overexpressed genes in each subclass were subsequently mapped to gene ontology, and significantly over-represented gene ontology terms were identified based on hypergeometric distribution with a false discovery rate <0.1.

Data access

Annotated somatic variants and interactive variant analysis results are available online at www.ingenuity.com/acrg2012.

Competing interest statement


Acknowledgments

This work was funded by Asian Cancer Research Group, Inc. (ACRG), a not-for-profit organization formed by Eli Lilly, Merck, and Pfizer. We thank Stephen Friend and Gang Jin for initiating the establishment of ACRG. We are grateful to all the past and present members of ACRG, especially Kerry Blanchard, Yaron Turpaz, Jonathon Sedgwick, Greg Tucker-Kellogg, Xiang S. Ye, Gary Gilliland, Carolyn Buser, Peter Shaw, Neil Gibson, and Steven Adams. We also thank Ingenuity Systems for providing the Ingenuity Variant Analysis system to enable readers to analyze annotated somatic variants interactively online.

Author contributions: The study was initiated and designed by M.M., H.D., C.R., and J.L. WGS sequencing and SNV validation was done by X.L., G.W., Q.Z., and B.W. Data analysis was conducted by H.Z., Z.K., S.L., R.C., K.H., Y.L., and H.G. M.D.W., P.W., M.Q., I.H.W., and T.D.B. did the JAK1 functional studies. P.A.R. analyzed the JAK1 protein structure. Z.G. was the project manager. N.P.L., R.T.P., and S.T.F. coordinated the collection of specimens and clinical data. M.M., Z.K., and S.L. wrote the manuscript. Z.K., S.L., H.Z., P.A.R., T.D.B., M.D.W., P.W., M.M., and N.P.L. produced tables and figures. All the authors contributed to the final manuscript.

References


1432  Genome Research  www.genome.org


BMC Genomics 12: 375.