The most frequent chromosomal structural loss in hepatocellular carcinoma (HCC) is of the short arm of chromosome 8 (8p). Genes on the remaining homologous chromosome, however, are not recurrently mutated, and the identity of key 8p tumor-suppressor genes (TSG) is unknown. In this work, analysis of minimal commonly deleted 8p segments to identify candidate TSG implicated GATA4, a master transcription factor driver of hepatocyte epithelial lineage fate. In a murine model, liver-conditional deletion of 1 Gata4 allele to model the haploinsufficiency seen in HCC produced enlarged livers with a gene expression profile of persistent precursor proliferation and failed hepatocyte epithelial differentiation. HCC mimicked this gene expression profile, even in cases that were morphologically classified as well differentiated. HCC with intact chromosome 8p also featured GATA4 loss of function via GATA4 germline mutations that abrogated GATA4 interactions with a coactivator, MED12, or by inactivating mutations directly in GATA4 coactivators, including ARID1A. GATA4 reintroduction into GATA4-haploinsufficient HCC cells or ARID1A reintroduction into ARID1A-mutant/GATA4-intact HCC cells activated hundreds of hepatocyte genes and quenched the proliferative precursor program. Thus, disruption of GATA4-mediated transactivation in HCC suppresses hepatocyte epithelial differentiation to sustain replicative precursor phenotype.
The most frequent chromosomal structural loss in hepatocellular carcinoma (HCC) is of the short arm of chromosome 8 (8p). Genes on the remaining homologous chromosome, however, are not recurrently mutated, and the identity of key 8p tumor-suppressor genes (TSG) is unknown. In this work, analysis of minimal commonly deleted 8p segments to identify candidate TSG implicated GATA4, a master transcription factor driver of hepatocyte epithelial lineage fate. In a murine model, liver-conditional deletion of 1 Gata4 allele to model the haploinsufficiency seen in HCC produced enlarged livers with a gene expression profile of persistent precursor proliferation and failed hepatocyte epithelial differentiation. HCC mimicked this gene expression profile, even in cases that were morphologically classified as well differentiated. HCC with intact chromosome 8p also featured GATA4 loss of function via GATA4 germline mutations that abrogated GATA4 interactions with a coactivator, MED12, or by inactivating mutations directly in GATA4 coactivators, including ARID1A. GATA4 reintroduction into GATA4-haploinsufficient HCC cells or ARID1A reintroduction into ARID1A-mutant/GATA4-intact HCC cells activated hundreds of hepatocyte genes and quenched the proliferative precursor program. Thus, disruption of GATA4-mediated transactivation in HCC suppresses hepatocyte epithelial differentiation to sustain replicative precursor phenotype.

Introduction
Hepatocellular carcinoma (HCC) is the world’s second leading cause of cancer death. In the effort to understand the molecular mechanisms of HCC genesis, a relevant lead is that the most recurrent chromosomal structural loss in HCC is of chromosome 8 (8p), occurring in more than 60% of cases (1, 2). The frequency of this event suggests loss of one or more 8p genes is central to transformation. Of the approximately 500 genes on chromosome 8p, however, none are mutated at a high rate—presumably, complete loss of function by mutation of the remaining alleles might be deleterious, and haploinsufficiency of specific 8p genes is enough to contribute to transformation. The lack of mutations, however, complicates the task of identifying the key tumor-suppressor genes (TSGs) and pathways disrupted by 8p deletion, and the identity of this gene or genes is unknown. One clue is that occasionally, instead of the usual deletion of the entire chromosome 8p, there are smaller deletions targeting the segment from 8p22 to 8p23 (minimal commonly deleted regions), removing tens instead of hundreds of genes (3).

There have been attempts at functional interrogation of this more limited catalog of genes. In one study, FGL1 (also known as LFIRE or HFREP-I), which encodes for a fibrinogen (blood coagulation) family protein, was identified as a candidate 8p TSG because of its location at 8p22 and its lower expression in HCC versus adjacent noncancerous liver and because its further knockdown in HCC cells by antisense oligonucleotides increased proliferation, while its exogenous introduction had the opposite effect (4). Fgl1-knockout mice, however, have no liver phenotype (5). In other studies, DLC1 on chromosome 8p22, which encodes for a GTPase-activating protein, was implicated as a TSG based on a range of observations similar to those generated for FGL1 (6, 7). Again though, Dlc1-haploinsufficient mice have no overt phenotype, liver or otherwise, and the complete knockout is embryonic lethal, with neural tube and cardiac but not liver defects (8). Thus, neither Fgl1 nor Dlc1 have been found to have a usual central role in liver cell fate determination to warrant the frequency of haploinsufficiency observed in HCC.

The epidemiologic link between HCC and inflammatory insults such as viruses suggests participation in transformation of pathways regulating liver regeneration. In several organ systems, a hierarchy of quiescent stem cells, then highly proliferative committed precursors (transit amplifying cells), then nonproliferating specialized mature cells, is well chronicled as mediating tissue homeostasis. At one point, the liver was thought to contrast with these other organs in that mature epithelial cells (hepatocytes...
and cholangiocytes) might respond to mitotic triggers to mediate regeneration/homeostasis (9). Recent detailed molecular and cellular examination, however, has shown that a cellular hierarchy is present in the liver. Despite difficulties in separating liver cells by morphologic appearance, a cellular subset displays high expression of early hepatocyte differentiation-driving transcription factors Tbx3, Foxa2, and Sox9, low expression of hepatocyte epithelial-differentiation genes, a diploid genome, and replicative potential; it also mediates liver homeostasis by differentiating into cells with a typical hepatocyte epithelial-differentiation gene expression profile and polyploid genome (9–12). Several groups have found that HCC phenocopies this subset of liver precursor cells (10, 13–19). How this happens, however, is unknown. Here, we show that haploinsufficiency of a master transcription factor driver of hepatocyte differentiation, GATA4, located on chromosome 8p, and genetic alterations that compromise the transactivating function of GATA4 are a cause of HCC precursor phenotype. Interestingly, another highly recurrent structural deletion in HCC, that of chromosome 17p, also centers on a gene for a master transcription factor, Tp53, the master regulator of apoptosis. Thus, the master regulatory circuits for 2 major MYC-antagonizing, proliferation-terminating metazoan programs, epithelial differentiation and apoptosis, are genetically disrupted in HCC.

Results

Candidate 8p TSGs. Chromosome 8p was frequently deleted in our HCC series from Singapore (24 of 55, 44% of cases). Deletion of 8p is also the most frequent structural deletion identified by SNP array karyotyping in the large TCGA series of patients with HCC in the USA (245 of 360, 68%) (Figure 1A). One patient in the Singapore series had a smaller deleted 8p segment that contained only 14 genes (Figure 1B). Of these 14 genes, only GATA4 (master transcription factor driver of hepatocyte differentiation) and NEIL2 (1 of 4 DNA glycosylases) had gene expression characteristics expected of candidate TSGs — lower expression in HCC with 8p deletion than without and in HCC versus normal liver (Figure 1B). Karyotyping of 26 HCC cell lines identified several minimal commonly deleted chromosome 8p segments that again centered close to GATA4 (Figure 1C) (Cancer Cell Line Encyclopedia [CCLE], https://portals.broadinstitute.org/ccle/home). In multiple publicly available gene expression databases, GATA4 was consistently less expressed in HCC versus normal liver and in HCC with 8p deletion than without, and a strong correlation (Spearman’s correlation, R = 0.5382, P < 0.0001, n = 361) was observed between GATA4 copy number and GATA4 expression (Figure 1D and Supplemental Figure 1; supplemental material available online with this article; https://doi.org/10.1172/JCI93488DS1).

Liver-conditional GATA4 haploinsufficiency produced a phenotype of persistent precursor proliferation and failed hepatocyte differentiation. Thus, integrated copy number and gene expression analyses implicated NEIL2 and GATA4 as candidate key 8p TSGs. Neil2-knockout mice, however, have no overt phenotype (20). In contrast, Gata4-knockout mice have liver agenesis, demonstrating a critical role for GATA4 in liver development (21, 22). To model the somatically acquired liver GATA4 haploinsufficiency of HCC, we crossed Gata4<sup>fl/fl</sup> mice to albumin promoter creatinine recombinase (Alb-cre) mice to produce progeny with liver-conditional deletion of a single Gata4 allele (Gata4<sup>WT/</sup>) (Figure 2A). Compared with Gata4<sup>fl/fl</sup> mice, Gata4<sup>WT/</sup> mice demonstrated substantially decreased GATA4 protein levels only in the liver (Figure 2B). At formal phenotyping at 3 and 8 months, the Gata4<sup>WT/</sup> mice (n = 11) had strikingly enlarged livers on a normal diet (Figure 2, C and D, and Supplemental Figure 2). Markers of proliferation Ki67 and MYC, measured by immunohistochemistry and Western blot, were substantially increased in Gata4<sup>WT/</sup> (Gata4 haploinsufficient) versus Gata4<sup>fl/fl</sup> (Gata4 WT) livers (Figure 2, E and F, and Supplemental Figure 2).

RNA-sequencing was used to compare the gene expression profiles of Gata4<sup>WT</sup>-haploinsufficient versus WT livers. Of genes more highly expressed in Gata4<sup>WT</sup>-haploinsufficient livers, 25% (197 of 800) were associated with a precursor stage of liver development (Figure 2G and Supplemental Table 1) (23). Of genes less expressed in Gata4<sup>WT</sup>-haploinsufficient livers, more than a third (300 of 800, 38%) were late liver development specialized hepatocyte epithelial function genes, such as lipid metabolism genes, coagulation pathway genes, and genes that antagonize MYC to terminate proliferation (24–26). Key hepatocyte precursor marker genes Afp, Axin2, and Cd34 were also measured by quantitative real-time PCR (QRT-PCR), affirming their enrichment with Gata4 haploinsufficiency (Figure 2H).

To better understand the level of the differentiation continuum at which hepatocyte maturation was impeded by Gata4 haploinsufficiency, we also used QRT-PCR to measure expression levels of key transcription factor drivers that operate at early versus late stages of hepatocyte epithelial differentiation, identified by others during hepatocyte generation in vitro and in liver development; expression levels of commitment and early hepatocyte differentiation–driving transcription factors Hnf6, Hhex, and Tbx3 were preserved while those of late-differentiation driving transcription factors Hnf4a, Hlf, and Nr1h4 were significantly suppressed with Gata4 haploinsufficiency (Figure 2H) (17–19, 27–29). GATA4 is a “pioneer” transcription factor, known to initiate remodeling of compacted chromatin for subsequent gene activation (30). We therefore asked whether baseline differences in the chromatin state could explain preserved/high expression of transcription factor genes Hnf6, Hhex, and Tbx3, but repression of Hnf4a, Hlf, and Nr1h4. Epigenetic marks at these genes were therefore examined at the ultimate baseline, embryonic stem cells (ESCs). Hnf6, Hhex, and Tbx3 had open chromatin in ESCs, that is, low DNA CpG methylation and high histone 3 lysine 4 trimethylation (H3K4me3) levels (Supplemental Figure 3). In contrast, suppressed late-differentiation transcription factor genes Hnf4a, Hlf, and Nr1h4 had closed or compacted chromatin at this same baseline, with high CpG methylation and low H3K4me3 levels (Supplemental Figure 3).

Hundreds of hepatocyte epithelial-differentiation genes are also suppressed and precursor genes increased in HCC compared with normal liver. Gene expression of adjacent noncancerous liver tissue and HCC was then compared by microarray (n = 46 pairs). Approximately 2,486 genes were expressed (average expression at least 1.5-fold higher in nonmalignant liver) in HCC (n = 46) than in nonmalignant liver (n = 46) (Figure 3A). The top tissue expression association of these suppressed genes was in liver genes, that is, approximately 25% of these genes (600 of 2,486) were specialized hepatocyte genes, such as members of the cytochrome p450
Figure 1. GATA4 is a candidate TSG on chromosome 8p. (A) Chromosome gains and losses in HCC. HCC (n = 55 Singapore series, n = 330 TCGA series) karyotyped by SNP array. Blue plot, frequency of loss by chromosome location; red plot, frequency of gain. (B) Of the 14 genes in the smallest minimal deleted region observed in a patient with HCC, only 2 were less expressed in HCC with 8p deletion (del) than without (WT) and less expressed in HCC than in normal liver (N). Gene expression by microarray. Wilcoxon rank sum test, 2-tailed. (C) Chromosome 8 (chr8) copy number loss peaked in proximity to GATA4, while copy number gain peaked in proximity to MYC in multiple HCC cell lines (n = 27). Representative images for JHH2, JHH4, JHH5, and SNU475 are shown. (D) Correlation between GATA4 mRNA expression and GATA4 GISTIC score (https://software.broadinstitute.org/software/cprg/?q=node/31). Spearman’s rank correlation coefficient, 2-tailed. TCGA, n = 361.
family and lipid metabolism genes (Figure 3A and Supplemental Table 2) \( (P < 1 \times 10^{-46}, \) Benjamini corrected, DAVID gene ontology analysis; https://david.ncifcrf.gov/). This pattern of suppression of hundreds of specialized hepatocyte genes was seen even in HCCs that appeared to be well differentiated by light microscopy (American Joint Committee on Cancer histologic grade 1; ref. 31.) (Supplemental Figure 4). As in murine conditional liver Gata4 haploinsufficiency, the problem seemed to be with late-stage maturation, since expression of transcription factors that drive hepatocyte commitment and early differentiation (HNF6, HHEX, HNFIA) was
preserved, but that of late-differentiation driving transcription factors (CEBPD, HLF, NR1H4, NR2FI) was suppressed (Supplemental Figure 5). More than 400 genes suppressed in HCC versus noncancerous livers were also suppressed in Gata4-haploinsufficient versus WT livers (Figure 3B). Specialized hepatocyte epithelial genes constituted a large fraction of these genes commonly suppressed in both HCC and Gata4-haploinsufficient livers (Figure 3C).

To complement the above analyses, genes differentially expressed between nonmalignant livers and HCC were identified also by another method, unsupervised hierarchical clustering. Of genes with significantly higher gene expression in HCC versus normal liver (false discovery rate [FDR] < 0.01), 90 of 700 (13%) were liver precursor genes (significantly linked with early stages of liver development) while only 10 of 700 (~1%) were terminal-differentiation genes. In contrast, of genes significantly less expressed in HCC versus normal liver, only 8 of 700 (~1%) were precursor genes, while 175 of 700 (25%) were terminal-differentiation genes (Figure 4, A and B, and Supplemental Table 3). As observed earlier for Gata4-haploinsufficient livers (Supplemental Figure 3), genes highly expressed in HCC, including proliferation genes (MYC target genes), have open or “poised” chromatin at baseline in ESCs, with low CpG methylation and high H3K4me3 levels, while in contrast, genes less expressed in HCC, including hundreds of known hepatocyte epithelial-differentiation genes, begin at this same baseline with closed chromatin characterized by high CpG methylation and low H3K4me3 levels (Figure 4, C–F).

A cause-effect role for GATA4 in the regulation of these suppressed hepatocyte genes was demonstrated by their upregulation upon exogenous correction of GATA4 levels in the HCC cell line PLC, which contains an 8p deletion (Figure 5A and Supplemental Figure 6); GATA4 markedly increased expression of 329 of 600 (~55%) of the hepatocyte genes suppressed in HCC (Figure 5A). Of particular interest are hepatocyte genes that antagonize MYC function to terminate proliferation; the hepatocyte factors HNF4A and CEBPD have been shown to have such actions (17–19, 27–29). GATA4 introduction activated both HNF4A and CEBPD quantified by qRT-PCR (Figure 5B). As expected, this was accompanied by downregulation of the MYC protein, upregulation of the p27/CDKN1B protein (p27/CDKN1B is the cyclin-dependent kinase inhibitor that mediates cell-cycle exits by differentiation — Cdkn1b knockout mice have >20% increase in liver mass; refs. 32–34) (Figure 5C), and decreased cell growth without early apoptosis (Figure 5D).
Atypical HCC containing a rare germline GATA4 loss-of-function mutation. To evaluate whether GATA4 is also inactivated by mutation in HCC, we sequenced all GATA4-coding regions from HCC and paired noncancerous liver. Remarkably, we found an identical germline GATA4 missense mutation (GATA4 V267M) in 2 cases of HCC without 8p deletion (frequency 2 of 51, 4%; DNA from peripheral blood mononuclear cells was sequenced to evaluate the germline) (Figure 6A and Supplemental Figure 7). The amino acid altered by this recurrent mutation was highly conserved across species (Supplemental Figure 7G). Although this mutation has not been noted in the sequencing of the general population of Singapore, it has been noted as a rare, exclusively heterozy-
gous germline variant in ethnic Han Chinese in China (4 of 957, 0.4%, rs116781972), where it has been linked with congenital heart defects (Figure 6B) (35–37). Both HCC cases with the mutation were atypical compared with other HCC patients; both were female, even though HCC rates are 4 times higher in males than females (38), neither patient had hepatitis B or C or cirrhosis or other HCC risk factors, and neither HCC contained frequent chromosomal structural gains or deletions by SNP array karyotyping, contrasting with most other HCCs in the series (Figure 6C). The mutation affected only 1 GATA4 allele (variant allele frequency, ~50%) (Supplemental Figure 7 and Supplemental Table 4).

Unlike gene deletion, missense alteration is not unmistakably a loss-of-function event. When we used expression vectors to express similar amounts of WT GATA4 or GATA4 V267M protein in GATA4-haploinsufficient HCC cells (PLC), however, GATA4 V267M was less able than GATA4 to activate HNF4A and CEBPD, downregulate MYC, upregulate p27/CDKN1B, or slow HCC cell growth (Figure 7, A–C). We therefore looked for biochemical reasons for decreased transactivating function of GATA4 V267M. Decreased DNA binding was not the explanation, since GATA4 V267M bound GATA-response elements with affinity equal to that of WT GATA4 in DNA-binding assays (Figure 7D). Nor was GATA4 V267M mislocalized to the cytoplasm instead of the nucleus by Western blot of cell fractions (Supplemental Figure 8). Amino acid 267 is located between the zinc finger DNA-binding domains of GATA4, a region thought to mediate protein-protein interactions. Thus, we comprehensively analyzed the GATA4 V267M and WT GATA4 protein interactomes by immunoprecipitation followed by liquid chromatography tandem mass spectrometry (LC-MS/MS). The unbiased protein interactome analysis revealed a striking absence of the mediator complex from the GATA4 V267M versus the WT GATA4 protein interactomes (Figure 7E, Supplemental Figure 9–11, and Supplemental Table 5). This finding was confirmed also by immunoprecipitation-Western blot for mediator 12 (MED12), performed in triplicate (Figure 7F and Supplemental Figure 11).

Major coactivators for GATA4 are recurrently inactivated in HCC. GATA4 V267M loss of interactions with mediator complexes suggested that direct inactivation of key coactivators that mediate transcription activation by GATA4 could be another mechanism by which GATA4 function is impaired in HCC. Coactivators for GATA4 were identified by LC-MS/MS analysis of the GATA4 protein interactome (Figure 8A and Supplemental Table 5). Several of the identified coactivators were recurrently inactivated at a high rate in HCC (~40% of cases), observed in both the Singapore and TCGA series, with the highest inactivating mutation rate observed for AT-rich interactive domain 1A (ARID1A) (Figure 8B and Supplemental Figure 12). HNF1A, a differentiation-driving transcription factor found in the GATA4 protein interactome, was also frequently mutated in both the Singapore and TCGA series (Figure 8, A and B, Supplemental Figure 12, and Supplemental Table 5). Consistent with the idea that coactivator inactivation will negatively affect GATA4-mediated gene activation, there was similar suppression of key late-differentiation factors downstream of GATA4 in both GATA4-haploinsufficient and ARID1A/SMARCA2/ARID2/SMARCA4-deficient HCC; this suppression was greater.
HCC cases were hepatitis B sAg positive). In contrast, approximately 25% of HCC in the TCGA series from the USA contained CTNNB1 mutations (Supplemental Figure 14). There was no significant difference, however, in the rate of GATA4 deletion, ARID1A mutation or deletion, or TP53 mutation or deletion in the Singapore versus TCGA series.

Discussion

The liver replenishes hepatocytes consumed by metabolic work and insults such as viruses routinely and in controlled fashion. One control on this tissue homeostasis is WNT-CTNNB1 signaling, which initiates tissue precursor proliferation (10), and mutations that constitutively activate CTNNB1 are frequent in HCC in Europe and North America (20%–40% of cases). CTNNB1 mutations are less frequent in Asia and in patients with hepatitis B–associated HCC (39–44), possibly because hepatitis B viral X protein directly increases WNT/CTNNB1 signaling (45, 46). Other homeostatic controls terminate precursor proliferation once it is initiated; advancing differentiation (transition to specialized lineage fate) routinely terminates exponential proliferation of pre-hepatocyte master transcription factor GATA4 and/or GATA4 coactivators impaired activation of terminal hepatocyte epithelial genes that physiologically antagonize MYC to mediate precursor-to-hepatocyte transition (Figure 9).

CTNNB1 mutations were not detected in our HCC series. Mutations in CTNNB1 are known to be less frequent in HCC in Asian subjects and/or patients with hepatitis B than in HCC in patients of other ethnicities or those without hepatitis B (refs. 39–43; reviewed in ref. 44). To extend these observations, we used targeted next-generation sequencing to evaluate for CTNNB1 mutations in the Singapore HCC series and found no mutations (mean coverage 2,777, range 38–5,645; 56% of the Singapore HCC cases were hepatitis B sAg positive). In contrast, approximately 25% of HCC in the TCGA series from the USA contained CTNNB1 mutations (Supplemental Figure 14). There was no significant difference, however, in the rate of GATA4 deletion, ARID1A mutation or deletion, or TP53 mutation or deletion in the Singapore versus TCGA series.

Figure 6. Atypical HCC containing a rare germline GATA4 mutation. (A) GATA4 germline missense mutation (V267M) found in 2 of 51 patients (4%). Sanger and targeted deep sequencing of HCC, adjacent nonmalignant liver, and peripheral blood mononuclear cells (Supplemental Figure 7). (B) Frequency of this mutation in healthy individual populations from Singapore (Sing.) and China (34–36) ***P < 0.001, Fisher’s exact test. (C) HCC cases with GATA4 mutation had near-normal karyotype. Karyotype analyzed by SNP array and imaged by IGV. Pathstage, American Joint Committee on Cancer staging system. HepBsAg, hepatitis B surface antigen; Del, GATA4 locus deletion.
that mimicked that of HCC. Presumably only 1 allele of GATA4 is inactivated in HCC because some GATA4 is needed for existence along the GATA4-dependent liver differentiation continuum; Gata4 complete knockouts have liver agenesis (21, 22). GATA4 is a known “pioneer” transcription factor, that is, GATA4 has been shown to access compacted chromatin that is inaccessible to lesser transcription factors to initiate the remodeling needed for subsequent gene activation (30). It is notable, therefore, that transcription factor drivers of proliferation, hepatocyte commitment, and early differentiation (Myc, Hnf6, Hhex, Tbx3) that had preserved or even increased expression in Gata4-haploinsufficient livers and HCC versus normal livers had chromatin already poised or open at the earliest baseline of ESC, while in contrast, transcription factor drivers of terminal epithelial differentiation (Hnf4a, Hlf, and Nr1h4) and hundreds of hepatocyte epithelial-differentiation genes that had significantly decreased expression in Gata4-haploinsufficient livers and HCC versus normal livers had closed chromatin at this same baseline. That is, marked baseline differences in the chromatin state of proliferation/commitment versus terminal differentiation genes may interact with pioneer/master transcription factor deficiency to stall differentiation at committed, intrinsically proliferative points in the differentiation continuum.

Also implicating chromatin remodeling failure in HCC genesis is direct inactivation of coactivators (chromatin remodelers) themselves — transcription factors are selective in their coactivator cursors, and degree of differentiation failure correlates with clinical aggressiveness of HCC (47). Even if differentiation derangement is not obvious by light microscopy, as can be the case with HCC classified as well differentiated (grade 1 histologic grade), it is evident by gene expression analyses that precursor genes are elevated and hundreds of hepatocyte genes are suppressed in grade 1 HCC compared with normal liver. That HCC phenocopies proliferating precursors with suspended epithelial differentiation has also been observed by others by (i) detailed immunophenotyping (15); (ii) a near-diploid genome characteristic of liver precursors (instead of the polyploid genome of mature hepatocytes) (10, 16); (iii) capacity of HCC cells to differentiate into hepatocytes (13, 14); and (iv) pericentral vein location of HCC that is similar to that of liver precursors (10, 16).

The mechanisms underlying this precursor phenotype of HCC were not known. A small subset of the hundreds of transcription factors expressed in cells are masters that command and coordinate other transcription factors to drive lineage fates. It is thus notable that a master transcription factor driver of hepatocyte lineage fate, GATA4 (24–26), is located in a minimal deleted segment of chromosome 8p in this and other HCC series (1, 3, 6, 7). Conditional deletion of 1 allele of Gata4 from murine liver, to model the haploinsufficiency observed in human HCC, created an enlarged, proliferative liver phenotype, with a gene expression profile of precursors and failed terminal hepatocyte epithelial differentiation that mimicked that of HCC. Presumably only 1 allele of GATA4 is inactivated in HCC because some GATA4 is needed for existence along the GATA4-dependent liver differentiation continuum; Gata4 complete knockouts have liver agenesis (21, 22).

GATA4 is a known “pioneer” transcription factor, that is, GATA4 has been shown to access compacted chromatin that is inaccessible to lesser transcription factors to initiate the remodeling needed for subsequent gene activation (30). It is notable, therefore, that transcription factor drivers of proliferation, hepatocyte commitment, and early differentiation (Myc, Hnf6, Hhex, Tbx3) that had preserved or even increased expression in Gata4-haploinsufficient livers and HCC versus normal livers had chromatin already poised or open at the earliest baseline of ESC, while in contrast, transcription factor drivers of terminal epithelial differentiation (Hnf4a, Hlf, and Nr1h4) and hundreds of hepatocyte epithelial-differentiation genes that had significantly decreased expression in Gata4-haploinsufficient livers and HCC had closed chromatin at this same baseline. That is, marked baseline differences in the chromatin state of proliferation/commitment versus terminal differentiation genes may interact with pioneer/master transcription factor deficiency to stall differentiation at committed, intrinsically proliferative points in the differentiation continuum.

Also implicating chromatin remodeling failure in HCC genesis is direct inactivation of coactivators (chromatin remodelers) themselves — transcription factors are selective in their coactivator...
female and did not have hepatitis B, alcohol use, or cirrhosis, and interestingly, neither HCC contained frequent chromosomal structural gains or deletions by SNP array karyotyping. Based on this atypical character of the cases and the enrichment of the mutation in the Singapore HCC series (4% from 0.4% reported in the general ethnic Han Chinese population), this germline mutation could be a risk factor for HCC.

Biochemical actions of coactivators to open chromatin are balanced by corepressors that keep it compacted. A corollary of coactivator loss of function, therefore, is unbalanced activity of opposing corepressors (52). This suggests pharmacologic inhibition of corepressors can potentially compensate for genetic reduction in coactivators to crack the chromatin barrier and resume epithelial differentiation of HCC cells (52). This result is observed in vitro with the clinical drug decitabine, which inhibits a corepressor, DNA methyltransferase 1 (DNMT1), that creates and/or maintains the CpG methylation repression marks strikingly enriched at terminal differentiation versus proliferation/commitment genes (48, 49) — and of the coactivators interacting with GATA4, identified by MS analysis of the GATA4 protein interactome, several (ARID1A, SMARCAD1, ARID2, SMARCA4) are genetically inactivated at a high rate in HCC (>40% inactivating mutations or deletions). Consistent with these loss-of-function events negatively affecting transactivation by GATA4, reintroduction of ARID1A into ARID1A-deficient/GATA4-intact HCC cells activated GATA4 target genes, resumed hepatocyte epithelial differentiation, and terminated proliferation. Also, liver-conditional deletion of 1 Arid1a allele in mice created a phenotype of proliferative liver precursors with impaired hepatocyte differentiation (50), and genetic inactivation of Smarcb1 in mice potently suppressed hepatocyte differentiation and increased proliferation (51).

We also found 2 HCC cases without GATA4 deletion, but with a very rare germline missense mutation in GATA4 (GATA4 V267M) that abrogated its interactions with mediator (MEDI2, etc.), another complex that mediates transcription activation. Both patients did not have usual HCC risk factors; they were of usage (48, 49) — and of the coactivators interacting with GATA4, identified by MS analysis of the GATA4 protein interactome, several (ARID1A, SMARCAD1, ARID2, SMARCA4) are genetically inactivated at a high rate in HCC (>40% inactivating mutations or deletions). Consistent with these loss-of-function events negatively affecting transactivation by GATA4, reintroduction of ARID1A into ARID1A-deficient/GATA4-intact HCC cells activated GATA4 target genes, resumed hepatocyte epithelial differentiation, and terminated proliferation. Also, liver-conditional deletion of 1 Arid1a allele in mice created a phenotype of proliferative liver precursors with impaired hepatocyte differentiation (50), and genetic inactivation of Smarcb1 in mice potently suppressed hepatocyte differentiation and increased proliferation (51).

We also found 2 HCC cases without GATA4 deletion, but with a very rare germline missense mutation in GATA4 (GATA4 V267M) that abrogated its interactions with mediator (MEDI2, etc.), another complex that mediates transcription activation. Both patients did not have usual HCC risk factors; they were female and did not have hepatitis B, alcohol use, or cirrhosis, and interestingly, neither HCC contained frequent chromosomal structural gains or deletions by SNP array karyotyping. Based on this atypical character of the cases and the enrichment of the mutation in the Singapore HCC series (4% from 0.4% reported in the general ethnic Han Chinese population), this germline mutation could be a risk factor for HCC.

Biochemical actions of coactivators to open chromatin are balanced by corepressors that keep it compacted. A corollary of coactivator loss of function, therefore, is unbalanced activity of opposing corepressors (52). This suggests pharmacologic inhibition of corepressors can potentially compensate for genetic reduction in coactivators to crack the chromatin barrier and resume epithelial differentiation of HCC cells (52). This result is observed in vitro with the clinical drug decitabine, which inhibits a corepressor, DNA methyltransferase 1 (DNMT1), that creates and/or maintains the CpG methylation repression marks strikingly enriched at terminal differentiation versus proliferation/commitment genes (48, 49) — and of the coactivators interacting with GATA4, identified by MS analysis of the GATA4 protein interactome, several (ARID1A, SMARCAD1, ARID2, SMARCA4) are genetically inactivated at a high rate in HCC (>40% inactivating mutations or deletions). Consistent with these loss-of-function events negatively affecting transactivation by GATA4, reintroduction of ARID1A into ARID1A-deficient/GATA4-intact HCC cells activated GATA4 target genes, resumed hepatocyte epithelial differentiation, and terminated proliferation. Also, liver-conditional deletion of 1 Arid1a allele in mice created a phenotype of proliferative liver precursors with impaired hepatocyte differentiation (50), and genetic inactivation of Smarcb1 in mice potently suppressed hepatocyte differentiation and increased proliferation (51).

We also found 2 HCC cases without GATA4 deletion, but with a very rare germline missense mutation in GATA4 (GATA4 V267M) that abrogated its interactions with mediator (MEDI2, etc.), another complex that mediates transcription activation. Both patients did not have usual HCC risk factors; they were female and did not have hepatitis B, alcohol use, or cirrhosis, and interestingly, neither HCC contained frequent chromosomal structural gains or deletions by SNP array karyotyping. Based on this atypical character of the cases and the enrichment of the mutation in the Singapore HCC series (4% from 0.4% reported in the general ethnic Han Chinese population), this germline mutation could be a risk factor for HCC.

Biochemical actions of coactivators to open chromatin are balanced by corepressors that keep it compacted. A corollary of coactivator loss of function, therefore, is unbalanced activity of opposing corepressors (52). This suggests pharmacologic inhibition of corepressors can potentially compensate for genetic reduction in coactivators to crack the chromatin barrier and resume epithelial differentiation of HCC cells (52). This result is observed in vitro with the clinical drug decitabine, which inhibits a corepressor, DNA methyltransferase 1 (DNMT1), that creates and/or maintains the CpG methylation repression marks strikingly enriched at terminal differentiation versus proliferation/commitment genes.
Figure 9. The model: several genetic alterations in HCC impair master transcription factor-mediated (GATA4) chromatin remodeling/transactivation to suppress precursor-to-epithelial transition. Normal liver cells have intact master transcription factors (e.g., GATA4/FOXA1) that dictate cell fate by recruiting coactivators (e.g., ARID1A) to remodel and activate downstream transcription factors and hundreds of hepatocyte epithelial-differentiation genes, antagonizing MYC in the process for terminal epithelial differentiation. In HCC, haploinsufficiency or loss-of-function mutation in 1 allele of GATA4 and/or inactivation of GATA4 coactivators (e.g., ARID1A) impairs this chromatin remodeling and transcription activation to produce a selective suppression of terminal epithelial-differentiation genes, a selective repression that is facilitated by the inherently closed chromatin state of these versus commitment and MYC proliferation genes.

Also, the chromatin landscape of ESCs, the ultimate tissue precursors, suggests that pioneer function is broadly less needed for activation of intrinsically accessible proliferation/commitment genes, but necessary for activation of intrinsically compacted terminal epithelial-differentiation genes. Notably, therefore, chromosome 8p deletion is also highly recurrent in cancers of the lung, colon, breast, prostate, head, neck, bladder, ovary, prostate, and brain, and minimal commonly deleted regions in these cancers also incorporate GATA4 (71–75). GATA4 reintroduction into glioblastoma cell lines promoted differentiation and decreased proliferation (71). It is conceivable, therefore, that GATA4 haploinsufficiency and/or coactivator inactivation is how MYC-antagonizing terminal-differentiation programs are selectively suppressed in multiple cancers.

Missing from our understanding of HCC genesis was the gene(s) and pathways driving selection pressure for highly recurrent chromosome 8p deletions. Here, we showed that GATA4 is a key TSG targeted by this alteration and that GATA4 loss of function explains in large part how HCC cells suspend the epithelial fate that would otherwise suppress replicative precursor phenotype.

Methods

Primary HCC and paired nonmalignant liver tissue. Fresh primary HCC tumors were surgically removed (therapeutic segmentectomy or hemihepatectomy) at the National Cancer Centre Singapore between 2008 and 2011. Resected liver tissue was analyzed by frozen section to identify neoplastic and nonneoplastic areas. These were macrdissected for analyses as HCC and paired nonmalignant liver tissue.

Liver-conditional knockout of Gata4. Gata4 double-floxed mice were purchased from the Jackson Laboratory (stock 008194). These mice were crossed to albumin-cre mice, also from the Jackson Laboratory (stock 003574). 100% of F1 progeny were Gata4 haploinsufficient only in hepatocyte cells. Both double-floxed and liver-conditional Gata4-haploinsufficient mice were monitored daily; animals with signs of distress were euthanized by an IUCAC-approved protocol. PCR genotyping primer (Supplemental Table 6) sequences were provided by the Jackson Laboratory. DNA was isolated from mouse tails using the DNA Purification Kit (Promega catalog A1020).

RNA and DNA extraction. Snap-frozen specimens were equilibrated with a buffer (RNAlater-ICE, Ambion) that preserves RNA integrity. Frozen specimens, no larger than 0.5 cm, were added to RNAlater-ICE that was first cooled to ~8°C in polypropylene tubes and equilibrated with the RNAlater-ICE at ~20°C overnight. The specimens were separated for (a) tissue homogenization and RNA extraction and (b) proteinase K digestion and DNA extraction. The fragment for RNA extraction was homogenized using a tissue homogenizer. The mirVana Kit (Ambion) was used per the manufacturer’s protocol for extraction of total RNA. The fragment for DNA extraction was minced with a blade into fragments approximately 2 mm in dimension. These fragments were placed in Eppendorf tubes with cold PBS and left on ice for 5 to 10 minutes to leach out the RNAlater-ICE before digestion in proteinase K for DNA extraction. After removing PBS, lysis buffer was added followed by proteinase K digestion as per the manufacturer’s instructions for the DNasey Blood and Tissue Kit (QIAGEN).

Cell culture and transfaction. Human HCC cell lines HepG2 (ATCC, stock HB-8065) and PLC (ATCC, stock CRL-8024) were cultured in RPMI media with 10% FBS, 100 U/ml penicillin, and 100 µg/ml strep-
tomycin (Mediatech). Cells were incubated at 37°C in a 5% CO₂ atmosphere. DNA was isolated from both cell lines for SNP array analysis. PLC cells were transfected with GATA4 WT and GATA4 V267M vector, and approximately 60 × 10⁶ cell pellets were harvested at 72 hours. Harvested cells were resuspended in 1× PBS plus PI plus PMSF (PBWS buffer PH7-9). The pellet was centrifuged at 835 g for 5 minutes. Cells were washed 3 times in PBWS buffer.

**GATA4 and ARID1A expression vectors.** WT GATA4 cDNA was cloned into pFlag-CMV4 (OriGene Technologies) using Clontech infusion cloning (Clontech Laboratories Inc.). The primers used for infusion cloning (Supplemental Table 6) were designed to insert WT GATA4 cDNA at the C terminal end of the Flag tag in the pFlagCMV4 vector (OriGene Technologies). Transient transfection of GATA4 WT or empty vector into HCC cell lines was performed using Xfect transfection (Clontech) following the manufacturer’s guidelines. Cell pellets were isolated at 0, 48, and 96 hours after transfection for downstream analysis. pLenti-puro-ARIDIA was a gift from Ie-Ming Shih (Johns Hopkins, Baltimore, MD, USA) (Addgene plasmid 39478) (76).

**In vitro site-directed mutagenesis.** In vitro site-directed mutagenesis was performed using the Stratagene QuickChange Site-Directed Mutagenesis Kit (catalog 200519). Mutagenic oligonucleotides specific for point mutation 11607635G>A in GATA4 exon 4 (Supplemental Table 6) were designed and ordered from Integrated DNA Technologies (ITD). Both forward and reverse primers annealed to the same sequence on opposite strands of the plasmid expressing GATA4 cDNA. The mutant strand synthesis reaction was performed with the recommended PCR conditions from a Stratagene kit and using Pfu Turbo DNA polymerase (Stratagene 200519). After the thermocycler reaction, the PCR product was treated with Dpn I endonuclease reagent to digest the WT parenteral DNA template. The Dpn I-treated PCR product was purified and used to transform XL1 Blue Supercompetent Cells. The cells containing the mutant CDNA were used to transform E. coli at 37°C overnight. Colonies were selected and mini-prep reaction was carried out using the QIAGEN Mini-Prep Kit (catalog 27106). The isolated bacteria DNA was sequenced to confirm the presence of the mutant GATA4 cDNA by both Sanger sequencing and target deep sequencing using primers designed to amplify WT GATA4 cDNA (data not shown) (Supplemental Table 2). Colonies with mutant GATA4 were mix-prepared using the Promega Pure Yield Plasmid Kit (catalog A2492) and used for transfection experiments.

**QRT-PCR using SYBR green.** RNA was isolated using the RNeasy method (QIAGEN), and cDNA was prepared using the iScript cDNA Synthesis Kit (Bio-Rad). QRT-PCR was done using an ABI Prism 7500 Sequence Detection System (Applied Biosystems) and SYBR Premix Ex Taq II (TakaRa). Real-time PCR primers (Supplemental Table 6) were designed with PrimerQuest (https://www.idtdna.com/primerquest/Home/Index). The relative number of copies of mRNA (RQ) was calculated based on the average Ct values using the housekeeping gene GAPDH as internal control and baseline controls for relative expression. Results are shown as mean ± SD of 3 independent experiments.

**Sanger and targeted next-generation sequencing.** All coding region exons of GATA4 in genomic DNA from paired primary HCC and nonmalignant liver tissue were Sanger sequenced by ABI 3730x1 DNA analyzer (Applied Biosystems). Primers for bidirectional sequencing were designed using PrimerQuest (www.idtdna.com) (Supplemental Table 6). DNA sequences were analyzed using FinchTV DNA analysis software. In patients in whom mutations were identified, DNA isolated from peripheral blood mononuclear cells was similarly sequenced to determine whether the mutation was germline.

In addition, targeted next-generation sequencing was applied to exon 4 of GATA4 (to confirm the mutation identified by Sanger sequencing) and to all coding exons of HNF1A, ARID1A, SMARCA4, ARID2, CTNNBI, and TP53. Primers (Supplemental Table 6) were designed using PrimerQuest (www.idtdna.com). Each target exon was amplified by PCR using 100 ng/μl or more of primary DNA. The PCR amplification protocol included an initial denaturation step (94°C for 5 minutes), followed by 19 cycles of denaturation (94°C for 30 seconds, at 25% ramp cooling temperature), annealing, and extension (60°C for 3 minutes at 40% ramp cooling). After PCR amplification, subsequent purification and sequencing library preparation were according to Illumina pair-end library protocol. Briefly, DNA was purified using magnetic beads. The purified PCR products were end repaired to introduce sticky ends using end-repair enzyme (NEB catalog E6050S). Paired-end adapters were then ligated using T4 DNA Ligase (NEB catalog M0202S) to the amplified PCR fragments of about 250 bp in length. Nick fill reaction was performed using Bst DNA polymerase (NEB catalog MO374S). Library fragments and all PCR amplification were performed using HotStart Applied Biological Material (ABM) Tag DNA polymerase (ABM catalog G011). The library was subjected to deep sequencing on Illumina MiSeq using 2 × 300 paired-end sequencing following standard loading procedures. Coverage information for each sequenced gene for each sample is reported in Supplemental Table 4. Raw sequencing reads were trimmed of Illumina adapters, and low quality reads were discarded. Paired-end alignment was done using BWA 0.6 aligner with hg19 as the reference sequence. Variant extraction was done using the GATK pipeline (GATKv3.3). We prioritized variants after an elimination of sequencing/mapping errors (removal of low-quality reads, sequencing depth of at least 30 reads and 10 mutation reads, removal of variants with directional bias), removal of potential, common, benign polymorphisms (using publicly available databases of common and rare germline variants such as ExAC [http://exac.broadinstitute.org/], ESP6500 [http://evs.gs.washington.edu/EVS/], and 1000 Genomes [http://www.internationalgenome.org/]), and removal of variants with minor allele frequency of more than 0.001. Sequences were aligned to the reference genome using NovoAlign and were analyzed using Integrative Genome Viewer (IGV) software.

**RNA sequencing.** RNA was extracted from murine livers by the methods discussed above; however, the liver samples were also treated with DNaseI to remove potential DNA contamination. RNA integrity was measured by 1% 2D gel electrophoresis. Samples with 28S and 18S rRNA ≥ 3 mice per group) were sent for sequencing by ABM. RNA quality check was confirmed using Agilent 2100 Bioanalyzer, with all samples passing quality control. The samples were subjected to poly A enrichment, followed by fragmentation, first- and second-strand synthesis, adenylation of 3′ ends, adapter ligation, DNA fragmentation enrichment, and real-time PCR quantification. Cluster generation and sequencing were carried out in one run on NextSeq 5000 (Illumina, cluster generation and 2-channel sequencing), and Bcl files were converted to FastQ data immediately after the run. Over forty million paired-end reads for all 6 samples were recovered from the sequencing. Sequence reads were aligned to murine reference genome mm10 by Bowtie 2, and expression data were generated for analysis. Heat maps were generated using ArrayStarv3 (DNASTAR).
Gene expression analysis by microarray. The HumanHT-12 v3 gene expression microarray (Illumina) was used to analyze RNA from paired HCC and nonmalignant liver. The array evaluated over 25,000 annotated genes with over 48,000 probes designed using RefSeq (Build 36.2, Rel 22) and UniGene (Build 199). Microarray probe intensity values were subject to average normalization by GenomeStudio software to minimize the effects of variation from nonbiological factors and to calculate expression measures from the raw data. Expression measures of probe sets covering specific genes of interest were exported as a spreadsheet to the SAS System V9.2 (SAS Institute Inc.) for further statistical analysis. Only probe intensity data with detection \( P \) values of less than 0.05 (a statistical calculation that provides the probability that the signal from a given probe is greater than the average signal from the negative controls) were used in analyses of differences between groups of samples. Heat maps were generated by ArrayStarv3 (DNASTAR). All original microarray data were deposited in the NCBI's Gene Expression Omnibus (GEO GSE57958).

High-resolution molecular karyotyping by SNP array. The Human-660W-Quad v1.0 DNA BeadChip Kit was used for high-resolution molecular karyotyping of DNA isolated from primary HCC specimens and a control nonmalignant DNA. The BeadChip analyzed more than 660,000 individual loci. Genome Studio (Illumina), KaryoStudio (Illumina), and IgV (77) software were used to document large chromosome aberrations (e.g., >75 kb), to score these aberrations as loss, gain, or uniparental disomy (UPD), and for crossmatching these aberrations with information from public databases. Affymetrix SNP6 CEL files of 27 liver-derived CCLE samples were converted into A- and B-allele frequencies using crlmm of the Bioconductor R package oligo. log2 R ratios (LRR) were then computed as log2 of A+B divided by its median. Curves fitted to LRR were obtained using the ggplot2 function stat_smooth with its span parameter set to 0.3. The smooth curve displayed without its data was found similarly, but using data from all 27 CCLE liver cell lines combined.

Cell fractionation and nuclear protein extraction. Cells were resuspended in 500 \( \mu \)l PBSW followed by addition of 10% NP-40 (1:20 \( \mu \)l). This was incubated on ice for 1 minute followed by centrifugation at 835 \( g \) for 10 minutes. Supernatant containing the cytoplasmic fraction was transferred to a separate tube. Nuclear pellets were washed in PBSW buffer and centrifuged at 835 \( g \) for 10 minutes. Benzonase was added (1 \( \mu \)l), and this was incubated on ice for 90 minutes and vortexed every 30 minutes. 250 \( \mu \)l of nuclear extraction buffer E1 (250 \( \mu \)l PBSW + 2% NP40 + 500 mM [5 \( \mu \)l + 5 M NaCl [25 \( \mu \)l]]) was added, followed by a 10-minute incubation on ice; mixture was vortexed every 5 minutes. Sample was centrifuged at full speed for 10 minutes. The supernatant, which contained the cytoplasmic fraction, was transferred to a separate tube. 5 \( \mu \)l of 10% SDS was added to the remaining nuclear pellet. Nuclear protein was extracted again by adding 250 \( \mu \)l of nuclear extraction buffer E2 (250 \( \mu \)l PBSW + 1% NP40 + 500 mM [2.5 \( \mu \)l NaCl] to the remaining pellet. Sample was incubated on ice for 10 minutes and vortexed every 5 minutes as above. This was followed by centrifugation at full speed for 15 minutes. Supernatant was added to the nuclear protein extraction tube. Nuclear extraction buffer (200 \( \mu \)l) was added to the tube containing the remaining pellet. This was incubated on ice for 10 minutes with vortexing every 5 minutes. The sample was centrifuged at full speed for 15 minutes, and supernatant was added to the nuclear pellet tube. Concentration of the total protein extracted was measured using BCA.

Covalent binding of antibody to protein G beads. 25 mg (200 \( \mu \)l) of protein G–sepharose was washed twice with 1X PBS, followed by incubation with 200 \( \mu \)l of Flag antibody for 1 hour at room temperature and pressure. Antibody-bound protein G was incubated in 1% chicken egg albumin for 1 hour. This was washed twice with 1X PBS. 25 mg of dimethylpimelimidate was added to 1 ml of 300 mM NEM, followed by swirling for 30 minutes at room temperature and pressure. This was repeated twice. Glycine-HCl (PH3), was added, followed by spin down. This was washed 3x using 1X PBS. Samples were then washed \( \times 2 \) using nuclear extraction buffer.

Immunoprecipitation. 200 \( \mu \)l of nuclear protein lysate was precleared using protein G-sepharose (50% slurry). This was incubated at 4°C for approximately 60 minutes and spun for 10 minutes at 4°C. Supernatant was transferred to fresh tubes. 30 mg nuclear protein extracts (precleared lysate) was transferred to tubes with antibody-bound protein G beads and rocked gently at 4°C overnight. This mixture was washed 5 times with 1X PBS containing 1% NP-40. Samples were dried using Spin-Dry vacuum centrifugation at \(-100 \) on a SpeedVac vapor trap. Immunoprecipitation products were extracted from the protein G beads using Laemmli sample buffer.

Western blot analysis. Western blot was by standard methods: antibodies used were GATA4 (Abcam, catalog Ab124265), anti-Flag (Sigma-Aldrich, catalog F7425-2MG), c-MYC (Cell Signaling Technology, catalog 5605), p27/CDKN1B (Cell Signaling Technology, catalog 3833) and 
β-actin (Sigma-Aldrich, catalog a3854), histone 3 (Cell Signaling Technology, catalog 9715S), MED12 (Cell Signaling Technology, catalog 4529S,) SMARCA5/SNF2 (Cell Signaling Technology, catalog 13544S), and anti-V5 (Abcam, catalog Ab27671).

Protein identification by LC-MS/MS. Anti-Flag and isotype antibody immunoprecipitation products were separated by molecular weight using SDS-polyacrylamide gel electrophoresis and stained with colloidal Coomassie blue (GelCode Blue, Pierce Chemical). All separated proteins on the gel slices were excised (Supplemental Figure 9); proteins were reduced with dithiothreitol (Sigma-Aldrich, D0632, 10 mM), alkylated with iodoacetamide (Sigma-Aldrich, I149, 55 mM), and digested in situ with trypsin. Peptides were extracted from gel pieces 3 times using 60% acetonitrile and 0.1% formic acid/water. The dried tryptic peptide mixture was redissolved in 20 \( \mu \)l of 1% formic acid for MS analysis. Tryptic peptide mixtures were analyzed by online LC-MS/MS on an Orbitrap Mass Spectrometer (Thermo Fisher Scientific).

Database search and data validation. Mascot Daemon software (version 2.3.2; Matrix Science) was used to perform database searches using the Extract_msn.exe macro provided with Xcalibur (version 2.0 SR2; Thermo Fisher Scientific) to generate peaklists. The following parameters were set for creation of the peaklists: parent ions in the mass range of 400–4500, no grouping of MS/MS scans, and threshold at 1000. A peaklist was created for each analyzed fraction (i.e., gel slice), and individual Mascot (version 2.3.01) searches were performed for each fraction. The data were searched against Homo sapiens entries in the UniProt protein database (May 2015 release, 151,569 total sequences; http://www.uniprot.org/). Carbamidomethylation of cysteines was set as a fixed modification, and oxidation of methionine was set as a variable modification. Specificity of trypsin digestion was set for cleavage after Lys or Arg, and 2 missed trypsin cleavage sites were allowed. The mass tolerances in MS and MS/MS were set to 10 ppm and 0.6 Da, respectively, and the instrument setting was spec-
ified as ESI-Trap. To calculate the FDR, the search was performed using the decoy option in Mascot. The spectral FDR and protein FDR were 0.35\% ± 0.17\% and 4.28\% ± 1.99\%, respectively. A minimum Mascot ion score of 25 and peptide rank 1 were used for automatically accepting all peptide MS/MS spectra.

**LFQ.** Relative protein quantification was performed using spectral count-based label-free relative protein quantitation (LFQ). For each biological sample, data from the individual gel slices were combined. Statistical analysis was performed on all proteins identified, with average spectral counts of 2 or more for at least 1 of the 3 experiments. The spectral count data were normalized by total spectral counts of the bait protein (GATA4) in each sample to adjust for differences in overall protein levels among samples. Proteins were considered to have a significant difference in abundance if there was a difference of 2-fold or greater in normalized spectral counts between experiments and a \( P \) value of 0.01 or less using a 2-tailed \( t \) test. Spectral counts for all proteins and peptides identified are provided in Supplemental Table 5.

**GATA4 DNA-binding analysis.** A 5′-GATA probe (ATTACGTGAT-GTGTCG3′) and negative control probe (5′-ATTACGCCCCAT-GGTGC3′) (underlines indicate GATA4 response element versus scrambled sequence ‘CCCCC’) were designed and ordered from ITD. The probes were biotinylated with a biotin-labeling kit following the manufacturer’s guidelines (Thermo Fisher Scientific, catalog K0651). GATA4 WT and GATA4 V267M vector were transfected in HCC cell PLC, cell pellets from 6.0 × 10⁶ cells were isolated at 48 hours, and nuclear fractions were extracted using nuclear extraction buffer: (250 µl PBSW buffer PH7-9) + (2% NP40 + 500 mM [5 µl] + 5 M NaCl [25 µl]). Isolated nuclear fractions were incubated with the GATA probe and the scramble control probe at 4°C overnight. Incubated fractions were pulled down using streptavidin beads for 1 hour at room temperature.

**Bioinformatic and statistical analysis.** Proteins identified by label-free LC-MS/MS were analyzed by the Ingenuity Pathway Analysis Tool (IPA, Ingenuity Systems). The core analysis function included in IPA was used to interpret the data in the context of biological processes, pathways, and networks. Right-tailed Fisher’s exact test was used to determine a \( P \) value indicating that the probability of biological functions, canonical pathways, and diseases associated with the networks was not because of chance alone. Each protein identifier was mapped to its corresponding gene object in the Ingenuity Pathways Knowledge Base.

For classification of hepatocyte precursor and hepatocyte genes, a public gene expression database (GEO GSE13149) of genes expressed sequentially from early to late stages of liver development was analyzed to identify genes enriched at different stages of development (23). Comparative Marker Selection (V10) in GenePattern (Broad Institute, Massachusetts Institute of Technology, Boston, Massachusetts, USA) (78) was used to find genes whose expression correlated with phenotype. Statistical significance was determined by the 1000 permutations test and an FDR (Benjamini-Hochberg) cutoff of less than 0.01. In a separate approach to identifying genes that were differentially expressed between HCC and paired nonmalignant liver, the average expression of each gene across 46 HCC samples and, separately, across 46 adjacent nonmalignant livers was calculated. Genes with average expression values in HCC (\( n = 46 \)) that were less than 66% of the average expression value in nonmalignant liver (\( n = 46 \)) were identified. To determine tissue expression (UNIGENE EST QUARTILE) associations, gene lists were uploaded into DAVID (79) to provide a ranked representation of tissue expression associations that were most saturated or “enriched” with the input gene lists. For analysis of public ChIP sequencing (ChIP-Seq) data, aligned Hi-ESC H3K4me3 ChIP-Seq data (ENCF7775QSE) were imported, analyzed, and visualized using EaSeq and its suite of integrated tools; all values were normalized to reads per million per 1 kbp (80).

Wilcoxon rank sum and Student’s \( t \) tests were 2-sided and performed at the 0.05 significance level unless stated otherwise. SDs for each set of measurements were calculated and represented as \( y \) axis error bars on each graph. JMP Pro 10.0 or SAS statistical software was used to perform statistical analysis (SAS Institute Inc., http://www.jmp.com) including correlation analyses.

**Study approval.** Human samples were obtained in the context of clinically indicated surgery and with written informed consent from patients in accordance with the Declaration of Helsinki and using protocols approved by the SingHealth Institutional Review Board at the National Cancer Centre Singapore; human HCC tumors and paired normal samples were surgically removed at the National Cancer Centre Singapore between 2008 and 2011. Animal studies were conducted at the Cleveland Clinic using protocols approved by the Cleveland Clinic IACUC.

**Author contributions**

All authors reviewed and edited the manuscript. WHS validated sequencing experiments of GATA4 and in vitro functional effects of GATA4 transfection. XG validated GATA4 immunoprecipitation. EQ generated the Gata4 conditional KO mouse model. BP performed bioinformatics and next-generation sequence analysis. HM analyzed GATA4 sequences. JB performed IHC studies. JN, CLC, RB, LSK, JL, NC, and MT isolated DNA and RNA from paired normal versus HCC samples. ZH isolated and analyzed gDNA from HCC and adjacent normal tissue. KPN analyzed ChIP-Seq data. JM reviewed deep sequencing data and edited the manuscript. TR performed biostatistics and analyzed SNP data. AC, LLO, YMT, PCC, PC, CYC, and KHL isolated and characterized tissues from HCC patients. LY and EH validated IHC studies. HCT and YS conceived hypotheses, designed experiments, obtained funding, analyzed data, and wrote the manuscript.

**Acknowledgments**

HCT is supported by the National Medical Research Council (NMRC)/MOHIAFCAT2/001/2014, the Biomedical Research Council (BMRC) EDB-IAF311020, a National Cancer Centre Research Fund (NCCRF) pilot grant, a Terry Fox Foundation Pilot Grant, and block funding from the National Cancer Centre Singapore. YS is supported by case GI SPORE P50CA150964, Comprehensive Cancer Center grant P30 CA043703, and Cleveland Clinic HCC Center of Excellence grants. Financial support was also received from the NIH (RO1 CA204373).

Address correspondence to: Han Chong Toh, Division of Medical Oncology, National Cancer Centre Singapore, Singapore. Phone: 65.64368314; Email: toh.han.chong@singhealth.com.sg. Or to: Yogen Saunthararajah, Taussig Cancer Institute, 9500 Euclid Avenue R40, Cleveland, Ohio 44195, USA. Phone: 216.444.8170; Email: saunthy@ccf.org.
1. Xue W, et al. A cluster of cooperating tumor-suppressor gene candidates in chro-

2. Kan Z, et al. Whole-genome sequencing identifies frequent mutations in hepatocellular carci-


4. Yan J, et al. LFIRE-1/HFREP-1, a liver-specific GTPase-activating protein with tumor suppres-

5. Demechy V, et al. Targeted deletion of fibrinogen like protein 1 reveals a novel role in endo-


7. Yuan BZ, Miller MJ, Keck CL, Zimonjic DB, Thorgersson SS, Popescu NC. Cloning, charac-


15. Köhn-Gaone J, Gogoi-Tiwari J, Ramam GA, Olynyk JK, Timnitz-Parker JE. The role of liver progenitor cells during liver regeneration, fibro-
genesis, and carcinogenesis. Am J Physiol Gastroen-
terol Liver Physiol. 2016;310(3):G143–G149.


17. Lazarevich NL, et al. Progression of HCC in mice is associated with a downregulation in the expression of hepatocyte nuclear factors. Hepa-


20. Chakraborty A, et al. Neill-2 null mice accumulate oxidized DNA bases in the transcriptionally active sequences of the genome and are sus-


22. Molkentin JD, Lin Q, Duncan SA, Olson EN. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogen-


29. Parviz F, et al. Hepatocyte nuclear factor 4α controls the development of a hepatic epi-

30. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chro-
matin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell. 2001;9(2):279–289.

31. Amin MB, et al. eds. AICC Cancer Staging Manu-

32. Fero ML, et al. A syndrome of multilobar hyper-


