Endogenous Retrotransposition Activates Oncogenic Pathways in Hepatocellular Carcinoma

Ruchi Shukla,1,11 Kyle R. Upton,2,11 Martin Muñoz-Lopez,3 Daniel J. Gerhardt,2 Malcolm E. Fisher,1 Thu Nguyen,2 Paul M. Brennan,4 J. Kenneth Baillie,1 Agnese Collino,5 Serena Ghisletti,9 Shruti Sinha,6 Fabio Iannelli,9 Enrico Radaelli,6 Alexandre Dos Santos,7,8 Delphine Rapoud,7,8 Catherine Guettier,7,8 Didier Samuel,7,8 Gioacchino Natoli,5 Piero Carninci,9 Francesca D. Ciccarelli,5 Jose Luis Garcia-Perez,5 Jamila Faivre,7,8 and Geoffrey J. Faulkner1,2,10,*

1Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush EH25 9RG, UK
2Cancer Biology Program, Mater Medical Research Institute, South Brisbane QLD 4101, Australia
3Department of Human DNA Variability, Pfizer-University of Granada and Andalusian Government Center for Genomics and Oncology (GENYO), 18007 Granada, Spain
4Edinburgh Cancer Research Centre, The University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh EH4 2XR, UK
5Department of Experimental Oncology, European Institute of Oncology (IEO), Via Adamello 16, 20139 Milan, Italy
6DIVET, School of Veterinary Medicine, University of Milan, Via Celoria, 20133 Milan, Italy
7INSERM U785, Centre Hépatobiliaire, Villejuif 94800, France
8Université Paris-Sud, Faculté de Médecine, Villejuif 94800, France
9RIKEN Yokohama Institute, Omics Science Center, 1-7-22 Suehiro-chō, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
10School of Biomedical Sciences, University of Queensland, Brisbane QLD 4072, Australia
11These authors contributed equally to this work
*Correspondence: faulknergj@gmail.com
http://dx.doi.org/10.1016/j.cell.2013.02.032

SUMMARY

LINE-1 (L1) retrotransposons are mobile genetic elements comprising ~17% of the human genome. New L1 insertions can profoundly alter gene function and cause disease, though their significance in cancer remains unclear. Here, we applied enhanced retrotransposon capture sequencing (RC-seq) to 19 hepatocellular carcinoma (HCC) genomes and elucidated two archetypal L1-mediated mechanisms enabling tumorigenesis. In the first example, 4/19 (21.1%) donors presented germline retrotransposition events in the tumor suppressor mutated in colorectal cancers (MCC). MCC expression was ablated in each case, enabling oncogenic β-catenin/Wnt signaling. In the second example, suppression of tumorigenicity 18 (ST18) was activated by a tumor-specific L1 insertion. Experimental assays confirmed that the L1 interrupted a negative feedback loop by blocking ST18 repression of its enhancer. ST18 was also frequently amplified in HCC nodules from Mdr2−/− mice, supporting its assignment as a candidate liver oncogene. These proof-of-principle results substantiate L1-mediated retrotransposition as an important etiological factor in HCC.

INTRODUCTION

Liver cancer accounts for 9% of all cancer deaths worldwide and 12% in developing countries (Jemal et al., 2011). Pathological inspection indicates hepatocellular carcinoma (HCC) in ~80% of liver tumors, with infection by hepatitis B virus (HBV) and hepatitis C virus (HCV) being the most prevalent risk factors, followed by chronic alcoholism (Jemal et al., 2011; Perz et al., 2006; Tateishi and Omata, 2012). Although early detection and monitoring of patients with liver cirrhosis can substantially improve 5 year survival rates, progression to advanced HCC reduces average life expectancy to less than 8 months (Llovet et al., 2008). As for other cancers, genome and exome resequencing have elucidated molecular pathways frequently perturbed in HCC (Guichard et al., 2012; Tateishi and Omata, 2012; Totoki et al., 2011), potentially enabling therapeutic intervention informed by the mutational signature of a given tumor. The capacity to catalog the full spectrum of genetic aberrations occurring in HCC is therefore of critical importance.

LINE-1 (L1) retrotransposons are a major source of endogenous mutagenesis in humans (Burns and Boeke, 2012; Levin and Moran, 2011). These mobile genetic elements utilize a “copy-and-paste” mechanism to retrotranspose to new genomic loci, with such success in germ cells that 500,000 L1 copies comprise ~17% of the genome (Lander et al., 2001). Of these copies, only 80–100 are transposition competent, with distinct subsets of frequently active—or “hot”—L1s driving insertional mutagenesis in each individual genome (Beck et al.,...
Retrotransposon insertions can profoundly alter gene structure and expression (Cordaux and Batzer, 2009; Faulkner et al., 2008; Han et al., 2004; Levin and Moran, 2011) and have been found in nearly 100 cases of disease (Faulkner, 2011; Hancks and Kazazian, 2012). L1 activity is consequently suppressed in most somatic cells by methylation of a CpG island in the internal L1 promoter (Coufal et al., 2009; Swergold, 1990). By contrast, L1 is often hypomethylated in tumor cells, removing a key obstacle to retrotransposition (Levin and Moran, 2011).

Despite this failure to repress L1 transcription, only a handful of L1 insertions have been found in human tumors until very recently (Liu et al., 1997; Miki et al., 1992). High-throughput L1 integration site sequencing has since revealed 9 and 69 de novo L1 insertions, respectively, in lung and colorectal tumors (Iskow et al., 2010; Solyom et al., 2012), whereas cancer genome resequencing elucidated a further 183 tumor-specific L1 insertions in colorectal, ovarian, and prostate cancer (Lee et al., 2011). In this latter study, more than half of all insertions were found in a single colorectal tumor; the other individuals presented fewer than five tumor-specific L1 insertions on average. These data suggest L1 mobilization may be common in epithelial tumors, though the reasons for possible cell-of-origin restriction are currently unknown.

Tumor-specific L1 retrotransposition has not previously been observed in HCC. For several reasons it is, however, a logical cancer in which to expect L1 mobilization. First, HCC is epithelial in origin. Second, HBV and HCV infection are common in HCC; viruses can suppress host defense factors, such as APOBEC proteins, that control retrotransposon activation. APOBEC3G has been shown, for instance, to inhibit both HBV replication and endogenous retrotransposition (Esnault et al., 2005; Turelli et al., 2004). Third, liver inflammation precedes HCC and may, via cellular stress, stimulate retrotransposition (Fornace and Mitchell, 1986). Given these facts, we aimed to map L1 integration sites in HCC using retrotransposon capture sequencing (RC-seq) and assess their impact upon oncogenic and tumor suppressor pathways.

**RESULTS**

**Enhanced Retrotransposon Capture Sequencing**

To test the hypothesis that L1 mobilizes in HCC, we applied an updated RC-seq protocol to 19 HCC tumors and matched adjacent liver tissue that were confirmed positive for HBV or HCV infection (Table 1). An earlier RC-seq design (Baillie et al., 2011) was modified to incorporate multiplex liquid-phase sequence capture (Figure 1A) using a refined probe pool (Table S1 available online) and a reduced insert size of ~220 nt, which enabled high-confidence assembly of overlapping paired-end 150 nt reads (Figure 1B). This change simplified genomic alignment and, more importantly, enabled single-nucleotide resolution of retrotransposon integration sites (Figure 1C).

Table 1. Nonreference Genome Insertions Detected by RC-Seq

<table>
<thead>
<tr>
<th>Donor Gender</th>
<th>Virus Age</th>
<th>Germline Insertions</th>
<th>Private Germline Insertions</th>
<th>Validated Tumor-Specific Insertions</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 M</td>
<td>HCV 65</td>
<td>2,082</td>
<td>202</td>
<td>3</td>
</tr>
<tr>
<td>15 M</td>
<td>HBV 53</td>
<td>1,845</td>
<td>216</td>
<td>1</td>
</tr>
<tr>
<td>21 M</td>
<td>HCV 51</td>
<td>2,019</td>
<td>271</td>
<td>0</td>
</tr>
<tr>
<td>29 M</td>
<td>HCV 52</td>
<td>1,602</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>32 M</td>
<td>HBV 73</td>
<td>1,681</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>33 F</td>
<td>HCV 57</td>
<td>1,982</td>
<td>234</td>
<td>2</td>
</tr>
<tr>
<td>35 F</td>
<td>HCV 78</td>
<td>1,786</td>
<td>96</td>
<td>0</td>
</tr>
<tr>
<td>42 F</td>
<td>HCV 67</td>
<td>1,594</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>47 M</td>
<td>HBV 61</td>
<td>1,581</td>
<td>77</td>
<td>2</td>
</tr>
<tr>
<td>48 M</td>
<td>HBV 35</td>
<td>1,744</td>
<td>212</td>
<td>0</td>
</tr>
<tr>
<td>49 M</td>
<td>HCV 68</td>
<td>1,644</td>
<td>58</td>
<td>0</td>
</tr>
<tr>
<td>60 M</td>
<td>HCV 48</td>
<td>1,570</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>62 M</td>
<td>HBV 33</td>
<td>1,750</td>
<td>153</td>
<td>0</td>
</tr>
<tr>
<td>70 M</td>
<td>HCV 55</td>
<td>1,673</td>
<td>82</td>
<td>0</td>
</tr>
<tr>
<td>86 F</td>
<td>HBV 56</td>
<td>1,701</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>89 M</td>
<td>HBV 60</td>
<td>1,739</td>
<td>163</td>
<td>4</td>
</tr>
<tr>
<td>95 M</td>
<td>HBV 54</td>
<td>1,773</td>
<td>88</td>
<td>0</td>
</tr>
<tr>
<td>106 M</td>
<td>HBV 60</td>
<td>2,141</td>
<td>48</td>
<td>0</td>
</tr>
<tr>
<td>116 M</td>
<td>HBV 62</td>
<td>1,532</td>
<td>71</td>
<td>0</td>
</tr>
</tbody>
</table>

F, female; M, male. Please see Tables S2 and S3 for supporting data and details.

A total of 7,689 nonreference genome insertions were detected in 19 tumor (T) samples and 19 matched nontumor (NT) liver samples. Of these, we annotated 7,644 as putatively germline (Table S3) because of their presence in (1) databases of retrotransposon-induced polymorphisms (Beck et al., 2010; Ewing and Kazazian, 2010; Iskow et al., 2010; Wang et al., 2006), (2) pre-existing insertions annotated by pooled blood RC-seq (Baillie et al., 2011), (3) multiple individuals, or (4) nontumor liver. L1, Alu, SVA, and LTR-flanked retrotransposons comprised 13.5%, 81.8%, 4.3%, and 0.4% of germline insertions, respectively. As expected, L1-Ta and L1-pre-Ta (99.3%) and AluY (99.7%) were the main L1 and Alu families active in germ cells (Mills et al., 2007).

A total of 2,241 germline insertions were found in only one individual each (Table 1 and Table S3) and were not annotated by the aforementioned retrotransposon polymorphism databases, suggesting that these were private or rare mutations or, alternatively, had occurred in early development (Garcia-Perez et al., 2007; Kano et al., 2009). RC-seq detected 1,489 (66.4%) insertions at both their 5’ and 3’ ends, enabling us to model the characteristic sequence features of L1-mediated retrotransposition.

Frequent Retrotransposition in the Human Germline

A total of 7,689 nonreference genome insertions were detected in 19 tumor (T) samples and 19 matched nontumor (NT) liver samples. Of these, we annotated 7,644 as putatively germline (Table S3) because of their presence in (1) databases of retrotransposon-induced polymorphisms (Beck et al., 2010; Ewing and Kazazian, 2010; Iskow et al., 2010; Wang et al., 2006), (2) pre-existing insertions annotated by pooled blood RC-seq (Baillie et al., 2011), (3) multiple individuals, or (4) nontumor liver. L1, Alu, SVA, and LTR-flanked retrotransposons comprised 13.5%, 81.8%, 4.3%, and 0.4% of germline insertions, respectively. As expected, L1-Ta and L1-pre-Ta (99.3%) and AluY (99.7%) were the main L1 and Alu families active in germ cells (Mills et al., 2007).

A total of 2,241 germline insertions were found in only one individual each (Table 1 and Table S3) and were not annotated by the aforementioned retrotransposon polymorphism databases, suggesting that these were private or rare mutations or, alternatively, had occurred in early development (Garcia-Perez et al., 2007; Kano et al., 2009). RC-seq detected 1,489 (66.4%) insertions at both their 5’ and 3’ ends, enabling us to model the characteristic sequence features of L1-mediated retrotransposition.
Each individual genome contained on average 244 nonreference genome L1 insertions, a figure 60% and 80% higher, respectively, than recent L1 insertion site sequencing on cell lines (Ewing and Kazazian, 2010) and single cells (Evrony et al., 2012). Therefore, to assess the RC-seq false-positive rate, we randomly selected 200 germline insertions (173 Alu, 14 L1, 11 SVA, and 2 LTR) for site-specific PCR validation (Table S5). Of these, we confirmed 197 (98.5%). The remaining three insertions (2 SVA and 1 Alu) occurred in repetitive genomic regions and were detected by multiple unique reads in at least ten different samples each, indicating that these may have represented PCR false negatives. These comparisons and experiments together demonstrate the sensitive and accurate mapping of bona fide retrotransposition events by RC-seq and further highlight ongoing L1 retrotransposition in the global human population (Beck et al., 2010; Ewing and Kazazian, 2010; Huang et al., 2010; Iskow et al., 2010).

**Activation of β-Catenin/Wnt Signaling via L1-Mediated Ablation of MCC**

To assess the potential tumorigenic consequences of the identified nonreference genome insertions, we selected and validated, by insertion site PCR, 31 L1, Alu, and SVA insertions in genes generally implicated to play a causal role in cancer (Futreal et al., 2004) or specifically in HCC (Guichard et al., 2012), including L1 insertions in the proto-oncogene AKT and the tumor suppressor FHIT (Table S5). Quantitative RT-PCR indicated, however, that 28/31 of these germline insertions did not significantly perturb host gene expression in tumor or nontumor liver versus control liver from five unaffected individuals (data not shown).

Strikingly, the three remaining insertions all coincided with strong inhibition of the tumor suppressor mutated in colorectal cancers (MCC) (Higgins et al., 2007). MCC is expressed in liver (Senda et al., 1999) and regulates the oncogenic β-catenin/Wnt signaling pathway frequently activated in HCC (Fukuyama et al., 2008; Guichard et al., 2012; Totoki et al., 2011). In vitro experiments have established that siRNA knockdown of MCC mRNA dramatically increases β-catenin (CTNNB1) expression, whereas MCC overexpression inhibits cellular proliferation (Fukuyama et al., 2008; Matsumine et al., 1996). MCC is also an intriguing HCC candidate gene because of its genomic proximity to APC, a major tumor suppressor mutated in familial adenomatous polyposis preceding colorectal cancer (Groden et al., 1991; Kinzler et al., 1991). It is important to note that mutated APC occurs in <2% of HCC cases versus >60% of colorectal carcinomas (Guichard et al., 2012; Powell et al., 1992). We therefore hypothesized that germline retrotransposition events specifically inhibited MCC tumor suppressor function in liver. To test this prediction, we assessed the impact of each MCC mutation upon MCC, APC, and CTNNB1 expression.

Three germline retrotransposon insertions were found in MCC. The first of these, labeled MCC-L1-α, comprised a 5.3 kb L1-Ta oriented in sense to MCC in donors 70 and 95 (Figure 3A). Another L1-Ta, labeled MCC-L1-β, was full-length (6 kb), occurred at a different genomic position in donor 116, and was oriented antisense to MCC (Figure 3B). Finally, in donor 33, we found an AluY (MCC-Alu; Figure 3C) inserted in an...
ENCODE-delineated enhancer (Thurman et al., 2012). Insertion site PCR revealed that MCC-L1-a was heterozygous in donor 70 and homozygous (or possibly hemizygous) in donor 95, whereas MCC-L1-b and MCC-Alu were heterozygous in donor 116 and donor 33, respectively (Figure 3D).

An immunoblot indicated that MCC was dramatically less abundant in tumor and nontumor samples from all four donors compared with control liver tissue (Figure 4A). By contrast, CTNNB1 was expressed much more strongly in the affected donors than in controls (Figure 4A). This inverse relationship was consistent with MCC suppression of CTNNB1 through protein-protein interactions, as reported elsewhere (Fukuyama et al., 2008). As a corroborating example, immunohistochemistry performed on tumor and nontumor tissue from donor 116 confirmed cytoplasmic CTNNB1 accumulation (Figure S1), a strong indicator that the factors controlling CTNNB1 expression outside of the plasma membrane were absent and that many cells had entered a proliferative state (Nhieu et al., 1999).

Quantitative RT-PCR indicated that MCC transcription was severely reduced (p < 0.002–p < 0.002, t test, degrees of freedom [df] = 19) in all four tumors compared to normal liver (Figures 4B). MCC-L1-α and MCC-L1-β strongly suppressed MCC expression in donor 95 and donor 116’s nontumor liver, respectively (Figure 4B). MCC was also significantly downregulated in tumor versus nontumor in all four individuals (p < 0.0001, t test, df = 4) but only by 30% versus normal liver controls. By contrast, MCC-L1-α, the homozygous L1 insertion in donor 95, severely reduced MCC transcription in both tumor (~83%) and nontumor (~63%) samples compared with normal liver controls (Figure 4B), demonstrating that the primary effect of MCC-L1-α was on MCC rather than APC. These data in sum confirmed that (1) L1-mediated retrotransposition in MCC specifically repressed MCC and not APC and (2) CTNNB1 was strongly induced in all four affected individuals, indicating activation of a major HCC oncogenic pathway.

Therefore, MCC-L1-α, MCC-L1-β, and MCC-Alu were the primary enactors of MCC transcriptional inhibition, potentially assisted by other modifications to MCC or its upstream regulatory pathway.

Finally, we performed qRT-PCR to evaluate APC transcription coincident with mutated MCC. We found no significant differential transcription of APC in tumor or nontumor liver from the four affected donors versus normal liver controls (Figure S2). In donor 95, APC was downregulated significantly in tumor versus nontumor (p < 0.003, t test, df = 4) but only by 30% versus normal liver controls. By contrast, MCC-L1-α, the homozygous L1 insertion in donor 95, severely reduced MCC transcription in both tumor (~83%) and nontumor (~63%) samples compared with normal liver controls (Figure 4B), demonstrating that the primary effect of MCC-L1-α was on MCC rather than APC. These data in sum confirmed that (1) L1-mediated retrotransposition in MCC specifically repressed MCC and not APC and (2) CTNNB1 was strongly induced in all four affected individuals, indicating activation of a major HCC oncogenic pathway.

Somatic L1 Mobilization in HCC

Forty-five nonreference genome insertions were annotated as tumor specific. These consisted of 17 L1, 27 Alu, and 1 SVA. We first validated each L1 insertion with insertion site PCR, including capillary sequencing of their 5’ and 3’ ends (Table S6). All 17 L1s successfully amplified; 12 confirmed as tumor-specific, and 5 were found in both tumor and nontumor liver. Further examination of the tumor-specific set revealed uniform usage of the degenerate L1 endonuclease motif highlighted in Figure 2B. In
two examples, PCR amplification of the 5’ junction was repeatedly unsuccessful, preventing TSD characterization, an outcome possibly due to gross genomic abnormality at the L1 insertion site (Gilbert et al., 2002). Eight of the other integration sites incorporated TSDs, whereas the remaining two examples involved small genomic deletions 3’ of the insertion site and no TSD.

Somatic L1 mobilization occurred in donors 12, 15, 33, 47, and 89 (Table 1), with the latter individual presenting four insertions. Two L1 copies (chr11:60136439 and chrX:99180431) were greater than 5.3 kb in length, but no insertions were full-length. All 12 somatic L1 insertions were from the L1-Ta subfamily.

We next evaluated 13 Alu insertions and the single SVA insertion found only in tumor, using insertion site PCR. In all cases, amplification occurred in both tumor and adjacent liver DNA, indicating germline insertions. Our primary explanation for this result is that there are several thousand potentially active AluY copies in the genome, compared to fewer than 100 active L1s (Bennett et al., 2008; Brouha et al., 2003). As seen previously, the RC-seq read count per Alu is consequently 75% lower than for L1 (Bailie et al., 2011), making false-negatives in the nontumor control more likely for Alu than for L1. A secondary explanation is that chromosomal gain is very common in HCC (Guichard et al., 2012), increasing the probability that some germline insertions are detected in tumor but not in adjacent nontumor liver. A final possibility is that mutations in individual precancerous cells are clonally amplified in tumors and are called as tumor-specific by RC-seq and germline by insertion site PCR. However, this was unlikely, as we consistently observed strong PCR amplification in both tumor and nontumor liver in these cases. Consequently, RC-seq reliably identifies new L1, Alu, and SVA mobilization events but requires insertion site PCR to annotate tumor-specific insertions.

In recent work, we reported somatic L1 mobilization in the normal brain but did not evaluate other organs (Bailie et al., 2011). For the current study, somatic L1 insertions in nontumor liver were considered difficult to evaluate because of the frequent occurrence of chromosomal loss in tumors. In this scenario, germline L1 insertions may be deleted in tumor but retained in nontumor liver and called somatic events. Nonetheless, we identified 21 L1 insertions restricted to nontumor liver in the set putatively annotated as germline and as a proof-of-principle experiment selected an example (chr13:27423763) for insertion site PCR and capillary sequencing (Table S6). This 2.5 kb L1-Ta insertion was detected only in liver and, interestingly, had a long (127 nt) TSD (Figure S3). A germline L1 insertion deleted in tumor cells would reasonably be expected to be detected in the nontransformed cells (e.g., lymphocytes) infiltrating a tumor (Unitt et al., 2005). Therefore, this very likely represented a bona fide liver-specific somatic L1 insertion in the preneoplastic liver of donor 47. Consequently, hepatocytes, or their progenitor cells, may support limited somatic L1 mobilization, though the contribution of this activity to malignancy remains unclear.

L1 Hypomethylation Enables Tumor-Specific Mobilization

To assess whether L1 activity and L1 methylation state were correlated in HCC samples, we performed bisulphite conversion...
of gDNA and capillary sequenced the CpG island present in the canonical L1 promoter. Eight tumors (15T, 47T, 48T, 62T, 89T, 95T, 106T, and 116T) matched adjacent liver samples, and control liver samples were analyzed. In the tumor group, 54.8% of L1-promoter CpG dinucleotides were methylated, compared with 69.2% in nontumor liver, a strongly significant difference (p < 2.5 \times 10^{-18}, chi-square test, n = 8) (Figure 5A).

On average, all but one CpG was hypomethylated in tumor, with the remaining CpG being equally methylated in tumor and nontumor liver (Figure S4A). Hypomethylation was not observed in grouped adjacent nontumor liver tissue versus controls. As shown in Figure S4B, a subset of four individuals (donors 47, 89, 106, and 116) presented much stronger L1-promoter hypomethylation in their tumor (40.5%) versus nontumor liver (72.3%) samples compared with the remaining individuals (69.2% versus 66.1%). The three individuals with tumor-specific L1 insertions and L1 methylation data (donors 15, 47, and 89) yielded a strong correlation between L1 hypomethylation percentage and tumor-specific L1 insertion count (r = 0.97; n = 3). Donor 89 exhibited the strongest tumor-specific L1 hypomethylation and also had the most tumor-specific L1 insertions (Figure S4C).

As for (B), except observed at the L1 5′ UTR (p < 0.006). Please see Figure S3 for an example of a somatic L1 insertion in nontumor liver.

Hypomethylation of the L1 promoter enables transcription of full-length L1 mRNAs that are translated to form the L1 mobilization machinery (Ostertag and Kazazian, 2001a). We therefore used cDNA synthesized with L1-specific primers (Wissing et al., 2012) to quantify L1 expression levels by TaqMan qRT-PCR. In this analysis, we measured L1 mRNA levels using primers targeting L1 ORF2 (Figure 5B) and the L1 5′ UTR (Figure 5C). In both
cases, significant enrichment was observed in tumor and non-tumor versus normal controls (p < 0.003 for ORF2, p < 0.006 for 5' UTR, t tests, df = 22). Together, these data showed that L1 was activated and transcribed in HCC, coincident with hypomethylation of the L1 promoter.

**ST18 Activated by a Tumor-Specific L1 Insertion**

Tumor-specific L1 insertions were observed in six protein-coding genes (Table S6). Quantitative RT-PCR indicated that two of these genes (STXBP5L and SLC5A8) were not expressed in liver. The expression of three other genes was reduced 2-fold to 6-fold in tumor versus adjacent liver (p < 0.05, t test, df = 4), including a 3' UTR insertion in SLC2A1 and intronic insertions in PHGDH and EFHD1 (Figure S5). These examples resemble those seen in other cancers in which intragenic L1 insertions in tumors coincided with reduced host gene expression (Lee et al., 2012). To our knowledge, downregulation of SLC2A1, PHGDH, or EFHD1 has not previously been associated with cancer.

The remaining tumor-specific L1 insertion occurred in donor 47 and was associated with activation of the transcriptional repressor suppression of tumorigenicity 18 (ST18), a member of the MYT1 zinc-finger transcription factor family (Yee and Yu, 1998). Contrasting reports depict ST18 as a tumor suppressor and as an oncogene in different cancers (Jandrig et al., 2004; Steinbach et al., 2006). ST18 is, however, very poorly expressed in liver (Jandrig et al., 2004), making it unlikely to act as a tumor suppressor in this context. Ectopic host gene expression was an unusual consequence of an L1 insertion given that these events are usually repressive (Han et al., 2004). As such, we hypothesized that ST18 was a candidate liver oncogene activated via an unknown mechanism triggered by an intronic L1 insertion.

Initial data from RC-seq indicated a heavily 5' truncated, 410 bp L1-Ta arranged antisense to ST18 (Figure 6A). The integration site incorporated a 17 nt TSD, a degenerate L1 endonuclease motif (GC/AAAA), and a 5' inversion (Figure 6B). We then confirmed these features by PCR amplification and capillary sequencing of the L1 5' and 3' junctions, indicating a tumor-specific L1 insertion (Figure 6C). PCR on DNA extracted from three distinct biopsies taken from the same tumor detected

**Figure 6. A Tumor-Specific L1 Insertion Causes Induction of ST18**

(A) ST18 mutant allele: a 0.4 kb L1-Ta insertion antisense to ST18. Primers used for PCR validation (1,2) are indicated above the gene.

(B) L1 insertion, magnified view: RC-seq detected the L1 5' and 3' termini, indicating a 17 nt TSD and a 5' inversion.

(C) Insertion-site PCR validation: the L1 was detected only in 47T, whereas the empty site was found in both 47T and 47NT.

(D) qRT-PCR: ST18 was upregulated 4-fold in 47T versus 47NT (p < 0.005, two-tailed t test, df = 4). Data are presented as mean ± SD.

(E) ST18 immunoblot: ST18 (115 kDa) was enriched in 47T versus 47NT and normal liver controls.

(F) ST18 immunohistochemistry: accumulation of ST18 (brown) was observed in tumor nodules compared to surrounding non-tumor regions. Nuclei were stained with hematoxylin (blue).

(G) A palindromic sequence motif was bisected by the L1. Each 8 nt unit (a and b, light green) contained a subsequence 1 nt different to a PIT1 enhancer motif known to bind MYT1 (Rhodes et al., 1993). A second motif ~58 bp from the L1 integration site matched the consensus CEBPA binding motif (orange).

(H) ChIP followed by quantitative real-time PCR in Huh7 cells confirmed enrichment for ST18 bound to the putative ST18-enhancer element illustrated in (G), compared to GAPDH. Data from antibodies targeting both the N termini and C termini of ST18 are shown. Significance values were calculated using two-tailed t tests (df = 4). Data are presented as mean ± SD.

Please see Tables S6 and S7 and Figures S5 and S6 for further information regarding tumor-specific L1 insertions and additional ST18 characterization.
the L1 in all three regions, suggesting clonal amplification of tumor cells with the L1 mutant ST18.

As noted above, qRT-PCR indicated that ST18 expression was significantly increased in tumor versus adjacent nontumor liver (p < 0.005, t test, df = 4) (Figure 6D). To corroborate this result, we performed an immunoblot and immunohistochemistry with an anti-ST18 antibody and found ST18 was indeed ectopically expressed in donor 47 tumor (Figures 6E and 6F). Chromosomal gain and regional copy number variation (CNV) have previously been reported for chromosome 8q, the genomic region containing ST18 (Guichard et al., 2012). However, quantitative real-time PCR on gDNA indicated no ST18 CNV in donor 47 tumor. Thus, tumor cells containing the ST18 L1 mutation were clonally amplified without CNV of the ST18 locus, followed by ST18 transcriptional activation.

In response, we predicted that ST18 was activated by insertional mutagenesis of a cis-regulatory element proximal to the L1. In silico analysis of the L1 integration site indicated that it bisected a palindromic motif containing two 8 bp units differing by one nucleotide and separated by 3 bp (Figure 6G). The probability of a random insertion in this motif, even allowing for a mismatch in the palindrome and a generous gap of ≤11bp, was less than 1/1,000 (permutation test). Intriguingly, each unit was only one nucleotide different to a strong MYT1 binding motif found in the enhancer of PTT1 (Rhodes et al., 1993). Previous experiments predicted that these units would bind MYT1 with reduced efficiency (Jiang et al., 1996), though transcription factors incorporating two zinc-finger domains, as for MYT1, are known to greatly gain efficiency through binding tandem DNA motifs (Yee and Yu, 1998). The putative MYT1 binding site was proximal to a strong binding site for CEBPA, a transcription factor enriched in liver and known to bind active enhancers (Johnson et al., 1987).

Based on this computational analysis, we predicted that the L1 bisected an enhancer normally bound to the zinc fingers of the ST18 MYT1 domain. To test this experimentally, we performed chromatin immunoprecipitation (ChIP) of DNA bound to the ST18 protein in Huh7 cells, followed by PCR amplification of the putative ST18 enhancer. This assay confirmed that, absent an L1 insertion, ST18 was preferentially bound to its own enhancer (p < 0.0004, t test, df = 4) (Figure 6H). An L1 insertion in the ST18 binding site would reasonably be expected to displace this repressive mark from the enhancer. Thus, we experimentally validated a model of ST18 activation in which a negative feedback loop was interrupted by a tumor-specific L1 insertion.

Finally, in view of the clonal amplification of tumor cells containing ectopically expressed ST18, we engaged complementary in vitro and in vivo experimental models to assess ST18 oncogenic function in HCC. Although ST18 is poorly expressed in liver, we found it to be abundant in several liver cancer cell lines (Figure S6A). We then determined the frequency of ST18 CNV in an Mdr2−/− mouse model of inflammation-driven HCC. TaqMan quantitative real-time PCR detected ST18 amplification in 4/23 Mdr2−/− HCC nodules and no deletions (Table S7). A disproportionally high percentage of advanced tumors (75%) presented ST18 amplification. ST18 expression was also significantly higher in nodules with amplified ST18 compared with wild-type mouse liver (p < 0.0001, t test, df = 19) (Figure S6B). These experiments demonstrate concordance of frequent ST18 amplification and upregulation in human and mouse models of HCC, results consistent with ST18 functioning as a candidate liver oncogene.

**DISCUSSION**

The present study highlights endogenous L1-mediated retrotransposition in the germline and somatic cells of HCC patients. We report two archetypal mechanisms revealing MCC and ST18 as HCC candidate genes. MCC is, for the many reasons highlighted above, a highly plausible liver tumor suppressor. Four out of 19 individuals studied here, including two cases each of HBV and HCV infection, presented distinct germline L1 or Alu insertions contributing to MCC suppression in tumor and non-tumor liver tissue. Strong upregulation of CTNNB1 in all four donors was consistent with prior observations that CTNNB1 is inhibited by MCC (Fukuyama et al., 2008). It is also interesting that MCC-L1-α was homozygous in donor 95, and therefore, MCC was almost certainly downregulated in the liver of this patient prior to HBV infection, i.e., preceding viral challenge, cirrhosis, and tumorigenesis.

We also demonstrate that MCC transcriptional repression in all four affected donors was exclusive of APC. Mutated APC is common in colorectal cancer but rare in HCC (Guichard et al., 2012; Powell et al., 1992). Even in colon, MCC presents numerous properties of a tumor suppressor (Bouwmeester et al., 2004; Fukuyama et al., 2008; Kohonen-Corish et al., 2007; Matsumine et al., 1996). Indeed, a Sleeping Beauty transposon mutagenesis screen using a mouse model of colorectal cancer found specific mutations in MCC and APC at a 1:9 ratio (Starr et al., 2009). Very recently, exome resequencing identified sporadic MCC point mutations in HCC (Guichard et al., 2012). Thus, MCC has potential to act as a liver tumor suppressor independent of APC, and our results support this potentially pivotal line of enquiry.

Tumorigenic retrotransposition in somatic cells was first observed 20 years ago, coincidentally in the APC gene of an individual with colorectal cancer (Miki et al., 1992). High-throughput sequencing has since provided the means to test whether tumor-specific retrotransposition is a common feature of cancer. Our results indicate that L1 mobilization occurs in a minority of HCC tumors, adding to the list of epithelial cancers (lung, colon, ovarian, and prostate) known to support the phenomenon (Iskow et al., 2010; Lee et al., 2012; Miki et al., 1992; Solyom et al., 2012). Although transformed tumor cells, including liver cancer cell lines, support frequent transgenic L1 mobilization (Moran et al., 1996), it is unknown whether endogenous L1 activation precedes neoplastic transformation in vivo. For this reason, it was interesting that L1 transcription was found in liver tissue adjacent to tumors, in addition to an example of somatic L1 mobilization. Finally, in a small cohort of tumor-specific L1 insertions, we identified mobilization via TPRT, twin priming, and a third mechanism resulting in a small deletion and no TSD, as reported elsewhere (Gilbert et al., 2002). These observations highlight the multiple routes by which L1 mobilization alters the tumor cell genome.
The results presented here corroborate recent data generated via whole-genome sequencing of other cancers. As in our study, Lee et al. (2012) described tumor-specific L1 insertions bearing the hallmark features of TPRT and also found intragenic L1 insertions in differentially expressed genes (Lee et al., 2012). One distinct feature of the current study is our discovery that germline L1 and Alu insertions significantly perturb expression of genes relevant to HCC. This is in contrast to another advance, the discovery made in our study that occasional activation of host genes by tumor-specific L1 insertions is relevant to HCC. Another advance is our explanation for the frequent amplification of ST18 in HCC nodules taken from Mdr2−/− mice, and (5) induction of ST18 transcription and translation in tumor not seen in adjacent nontumor liver or control liver, (3) consistent ST18 expression in transformed liver cancer cell lines, (4) frequent amplification of ST18 in HCC nodules taken from Mdr2−/− mice, and (5) induction of ST18 transcription in those animals. However, we do not make any conclusion regarding the function of ST18 as a tumor suppressor or oncogene outside of the liver and draw attention to the possibility that L1 activity varies substantially between individuals and cancer types (Iskow et al., 2010; Lee et al., 2012; Solyom et al., 2012). It remains to be proven whether this phenomenon correlates with prognosis, is useful in a diagnostic capacity, or can be subjected to exogenous interference in vivo. Nonetheless, we can conclude that L1-mediated retrotransposition is a potentially crucial source of mutations that can reduce the tumor suppressive capacity of somatic cells in HCC.

**EXPERIMENTAL PROCEDURES**

Full protocols can be found in the Extended Experimental Procedures.

**Samples**

Tumor and nontumor liver tissues from 19 HCC patients with a confirmed HBV or HCV infection were provided by the Centre Hépatobiliaire, Paul-Brousse Hospital. DNA and RNA were extracted with a DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany) and a mirVana miRNA Isolation Kit (Life Technologies, Carlsbad, CA, USA), respectively. Control liver samples from five donors were provided by the Edinburgh Sudden Death Brain and Tissue Bank. DNA and RNA were isolated through standard phenol-chloroform extraction and RNA-Bee RNA isolation reagent (Tel-Test), respectively. Samples were analyzed with approval from the French Institute of Medical Research and Health (Ref: 11-047), the East of Scotland Research Ethics Service (Ref: LR/11/ES/0022), and the Mater Health Services Human Research Ethics Committee (Ref: 1915A).

**RC-Seq Library Preparation, Sequencing, and Analysis**

Multiplexed DNA sequencing libraries were constructed for HCC tumor and nontumor samples using a paired-end illumina TruSeq Kit with substantial modifications. Briefly, 1 μg of sonicated DNA size selected for an insert size of 200–250 bp was used for each library and amplified by six cycles of ligation-mediated PCR (LM-PCR). Libraries were then pooled in groups of 4 to 6 and hybridized to an updated custom Roche NimbleGen sequence capture array comprising oligos tiling the 5′′ and 3′ termini of active human retrotransposon consensus sequences (Figure 1; Table S1). Libraries were again amplified by six cycles of LM-PCR and sequenced on an Illumina Hiseq2000. After quality filtering, each read pair was assembled into a contig, aided by 2 × 150-bp sequencing and a 220 nt insert size. Read contigs were then aligned to retrotransposon consensus sequences to determine their retrotransposon donor family, aligned to the human reference genome (hg19) to determine their genomic position, and finally formed into clusters.

**PCR Validation**

Germline retrotransposon insertions detected by RC-seq were first validated by a standard empty sitetilled site PCR assay and then, if unsuccessful, with POR targeting an insertion site 5′′ or 3′′ end. Tumor-specific insertions were characterized with a similar strategy but also incorporated 5′′ and 3′′ end capillary sequencing. All validation was performed on nonamplified DNA stored and handled separately from postamplification RC-seq products. Primers were designed using custom Python scripts and Primer3.

**qRT-PCR**

Complementary DNA was synthesized from total RNA using random hexamers, except for L1 analyses, where a specific sense L1 primer was used. qRT-PCR was performed using a LightCycler 480 (Roche, Indianapolis, IN, USA), and values were normalized to TATA-binding protein (TBP). For primer sequences, see Table S8.

**ACCESSION NUMBERS**

RC-seq FASTQ files were deposited in the European Nucleotide Archive (ERP001476).

**SUPPLEMENTAL INFORMATION**

Supplemental information includes Extended Experimental Procedures, six figures, and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2013.02.032.

**ACKNOWLEDGMENTS**

We thank Professor Haig Kazazian, Professor Bert Vogelstein, Dr. Bruno Amati, and Dr. Alistair Funnel for helpful discussion. We thank Professor Daniel Azoulay, Professor Denis Castaing, Professor Rene Adam and Dr. Eric Vibert (Centre Hépatobiliaire, Villejuif), and the Tissue Biobank Group (AP-HP and Université Paris-Sud) for providing HCC specimens. We thank Dr. Christine Beck and Professor John Moran for providing the genomic coordinates of a recently published full-length L1 insertion cohort (Beck et al., 2010). Research performed in the laboratories of D.S., G.N., P.C., F.D.C., J.F., and G.J.F. was funded by the European Union’s Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 259743 underpinning the MODHEP consortium. J.K.B. and P.M.B. were supported by Wellcome Trust Clinical Fellowships (090385/Z/09/Z and 090386/Z/09/Z, respectively) through the Edinburgh Clinical Academic Track (ECAT). J.L.G.P was supported by the British BBSRC (BB/H005935/1) and a C.J. Martin Overseas Based Biomedical Fellowship from the Australian NHMRC (575585).
REFERENCES


