Supporting Information

Lock et al. 10.1073/pnas.1405507111

SI Text

1. Candidate Gene Function Synopsis

Microfibrillar Associated Protein 5. The microfibrillar associated protein 5 (MFAP5, also known as MAGP2) gene encodes a 25-kDa microfibril-associated glycoprotein (1) that can activate Notch signaling (2) and has been shown to negatively affect serous ovarian cancer patient survival (3). However, this gene or gene product has not previously been described in association with lymphoma.

Dermatan Sulfate Proteoglycan 3 Precursor. The dermatan sulfate proteoglycan 3 precursor (EPYC) gene encodes a member of the small leucine-rich repeat proteoglycan family (epiphycan) and regulates fibrillogenesis by interacting with collagen fibrils and other extracellular matrix proteins; hence, it is involved in bone formation and also in establishing the ordered structure of cartilage through matrix organization. Increased EPYC mRNA expression has been observed in patients with invasive pancreatic ductal adenocarcinoma, compared with normal pancreatic tissues (4), but has not previously been associated with blood cancers.

Protein Tyrosine Phosphatase, Receptor Type F. Protein tyrosine phosphatase, receptor type F (PTPRF, or LAR) protein is member of the protein tyrosine phosphatase family of signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, cell cycle, and oncogenic transformation. Two alternatively spliced transcript variants of this gene, which encode distinct proteins, have been reported. PTPRF possesses an intrinsic protein tyrosine phosphatase activity (5). PTPRF expression is significantly increased in thyroid carcinomas (6) and breast cancer (7).

Aldehyde Dehydrogenase 1 Family, Member L1. Aldehyde dehydrogenase 1 family, member L1 (ALDH1L1, or FDH) protein encoded by this gene belongs to the aldehyde dehydrogenase family and catalyzes the conversion of 10-formyltetrahydrofolate, $NADP^+$, and water to tetrahydrofolate, NADPH, and CO_2 (8). Alternative splicing results in multiple transcript variants (9). Loss of function or expression of ALDH1L1 is associated with decreased apoptosis, increased cell motility, and cancer progression. ALDH1L1 is expressed in normal adult liver and down-regulated in liver cancer (10).

Fatty Acid-Binding Protein 7. Brain fatty acid-binding protein (FABP7, or BLBP) is a member of the FABP family of lipid chaperones that are involved in the uptake, storage, and intracellular trafficking of fatty acids (11). FABP7 is expressed in normal brain and is up-regulated in various solid cancers, including glioma (12), melanoma (13), renal cell carcinoma (14), and aggressive triple negative breast cancer (15) and is associated with poor prognosis. FABP7 has not been associated with blood cancers before this report, to our knowledge, and the native FABP7 promoter is not functional in normal blood cells.

2. Transposable Element-Gene Chimeric Sequences from RACE or RT-PCR

LTR2-FABP7. SUDHL4 cell-line [diffuse large B-cell lymphoma (DLBCL)] RT-PCR sequence (GenBank accession no. KM111186)

ACTTCTTTCTTTGGAGGCAAAAATTGGGTATAAGAC-AATATGAGGGGTGGTCTCCTCCCTTAAACCTGGTGGA-**ATGCAACTGAAGGAGAAGGATTGGAGATCTGCACTTATCA-AAACAGCAAATCAACAAAAA**GGCGTGGGCTTTGCCACTA-GGCAGGTGGGAAATGTGACCAAACCAACGGTAATTA-TCAGTCAAGAAGGAGACAAAGTGGTCATCAGGACTCT-CAGCACATTCAAGAACACGGAGATTAGTTTCCAGCTG-GGAGAAGAGTTTGATGAAACCACTGCAGATGATAGAA-ACTGTAAGTCTGTTGTTAGCCTGGATGGAGACAAACTT-GTTCACATACAGAAATGGGATGGCAAAGAAACAAATT-TTGTAAGAGAAATTAAGGATGGCAAAATGGTTATGAC-CCTTACTTTTGGTGATGTGGTTGCTGTTCGCCACTATG-AGAAGGCATAA

In the red LTR2 sequence, light blue shows splice junctions, splicing into exon 2, then 3 and 4. In green, ATG is present in the alternative first exon. Bold type represents the translated portion of the first exon, specific to chimeric mRNA.

MER57B-ALDH1L1. Karpas 422 cell-line (DLBCL) RACE sequence (GenBank accession no. KM111187)

CTGGnGTCTCTCTGATCTGCTGTGATTCTGAGGGCT-GCCCAATTTGTGAATCGTTCATTGCTCAATTAAACTTT-TTTAAATTTAATTTGGCTGAAGTTTTTCTTTTAACATGG-TGTCAGAAGCGGGATCCAAAGTACAGCTTCTAGCGAC-CCCCAGGAGTACTGAGTGAACAAGCAAGGTCCTTCCA-ACCCTCCTGCTACC**ATG**AAGATTGCAGTGATTGGACAG-AGCCTGTTTGGCCAGGAAGTTTACTGCCACCTGAGGAA

The MER57B sequence is shown in red; light blue is the splice junction, splicing into exon 2.

IM9 cell line RACE sequence (GenBank accession no. KM111188)

TTTGTGAATCGTTGATTGCTCAATTAAACTTTTTTAA-ATTTAATTTGGCTGAAGTTTTTCTTTTAACATGGTGTC-AGAAGCGGGATCCAAAGTACAGCTTCTAGCGACCCCC-AGGAGTACTGAGTGAACAAGCAAGGTCCTTCCAACCC-TCCTGCTACC**ATG**AAGATTGCAGTGATTGGACAGAGC-CTGTTTGGCCAGGAAGTTTACTGCCACCTGAGGAAGG-AGGGCCACGAAGTGGTGGGTGTGTTCACTGTTCCAGA-CAAGGATGGAAAGGCCGACCCCCTGGGTCTGGAAGC-TGAGAAGGATGGAGTGCCGGTATTCAAGTACTCCCG-GTGGCGTGCAAAAGGACAGGCTTTGCCTGATGTGGT-GGCAAAATACCAGGCTTTGGGGGCCGAGCTCAACGTC-CTGCCCTTCTGCAGCCAATTCATCCCCATGGAGATAAT-CAGTGCCCCCCGGCATGGCTCCATCATCTATCACCCGT-CACTGCTCCCTAGGCACCGAGGGGCCTCGGCCATCA-ACTGGACCCTCATTCACGGAGATAAGAAAGGGGGGG-TTTTCCATCTTCTGGGCGGATGATGGTCTGGACACC-GGAGACCTGCTGCTGCAGAAGGAGTGTGAGGTGCTCC-CGGACGACACCGTGAGCACGCTGTACAACCGCTTCCT-CTTCCCTGAAGGCATCAAAGGGATGGTGCAGGCCGTG-AGGCTGATCGCTGAGGGCAAAGCCCCCAGACTCCCTC-AGCCTGAGGAAGGAGCCACCTATGAGGGGATTCAGA-AGAAGGAGACAGCCAAG

The MER57B sequence is shown in red; light blue shows splice junctions, splicing into the second exon, then the third, fourth, and fifth. ATG present in exon 2 is shown in green.

THE1A-MFAP5. NUDUL1 cell line (DLBCL) RT-PCR sequence (GenBank accession no. KM111189)

TAAGGTCTCCCCAGCCATGTGAAACTCCTCATTGTA-ACACACATTCTACGCCTAGCCTGGCTTTCTTGCTCTCCC-TCATCTCATTGTTTCAGCGGAGGCCAAATCTGAAGTCC-TTTCCAGGGAGTGGCTCTGTTCATCTTATTCGCCAGCC-AAAGTAGGAACAGCGTAAGAGGAGAGAGACACATTC-AGCAGCCAAAGGACTCGGTGGAAAGAGCAGAACACC-

ATAGACAATATGTCGCTCTTGGGACCCAAGGTGCTGC-TGTTTCTTGCTGCAT

The THE1A sequence is shown in red; light blue shows splice junctions, splicing into the eighth base of exon 1. ATG present in exon 2 is shown in green.

L2-EPYC. OCL-Ly3 cell line (DLBCL) RT-PCR sequence (GenBank accession no. KM111190)

GCCAACCAACCACCACTTTGAAAGGACTGCAGAGG-AAAACCTTTGTCAGGAGACAGCCAGATTCTTTTAAGG-AAAGCTTGAAAAATGAAGACATTAGCAGGACTTGTTC-TGGGACTTGTCATCTTTGATGCTGCTGTGACTGCCCCA-ACTCTAGAGTCCATCAACTATGACTCAGAAACATATGA-TGCCACCTTAGAAGACCTGGATAATTTGTACAACTATG-AAAACATACCTGTTGATAAAGTTGAGATTGAAATAGC-CACAGTGATGCCTTCAGGGAACAGAGAGCTCCTCACT-CCACCCCCACAGCCTGAGAAGGCCCAGGAAGAGGAAG-AGGAGGAGGAATCTACTCCCAGGCTGATTGATGGCTC-TTCTCCCCAGGAGCCTGAATTCACAGGGGTTCTGGGG-CCACACACAAATGAAGACTTTCCAACCTGTCTTTTGTG-TACTTGTATAAGTACCACCGTGTACTGTGATGACCATG-AACTTGATGCTATTCCTCCGCTGCCAAAGAA

The L2a sequence is shown in red; light blue shows splice junctions, splicing into exon 2. ATG present in exon 2 is shown in green.

LTR16A1-L2-PTPRF. L591 cell line (HL) RT-PCR sequence (GenBank accession no. KM111191)

AAGGAGACAGCCACGCAAATATCTGGGAAAAGAG-CTCCAGGTGCCATAGGAGGAGCGGCGTCAGAAGGG-ACTGATGACTGATTGGACAGTGGTGCAGGAGGGAAG-GTCTCTGCATGTCAGTGATTAAATTGCTGAAGATGCCA-CTGCCTGAGACGGGCAGTATTGAAGGAAGGAGTGGAG-GCCCTGGTGCCCGGCCCTTGGTGCTGAGTATCCAGCAA

L2 is shown in red; light blue shows splice junctions, splicing into an intronic region then into exon 2. ATG was present in exon 3 but not sequenced.

3. Native FABP7 Amino Acid Sequence Compared with Chimeric FABP7 Amino Acid Sequence

Native FABP7 amino acid sequence: MVEAFCATWKLTN-SQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGDKVVIR-TLSTFKNTEISFQLGEEFDETTADDRNCKSVVSLDGDKLVH-IQKWDGKETNFVREIKDGKMVMTLTFGDVVAVRHYEKA

Chimeric FABP7 predicted amino acid sequence derived from SUDHL-4 amplification and sequencing:

MQLKEKDWRSALIKTANQQKGVGFATRQVGNVTKP-TVIISQEGDKVVIRTLSTFKNTEISFQLGEEFDETTADDRN-CKSVVSLDGDKLVHIQKWDGKETNFVREIKDGKMVMT-LTFGDVVAVRHYEKA

In the underlined sequence, different first exons result in different amino acid sequences. Red highlighting shows that the conserved nuclear localization sequence [conserved K (lysine)] is missing in chimeric FABP7.

1. Penner AS, Rock MJ, Kielty CM, Shipley JM (2002) Microfibril-associated glycoprotein-2 interacts with fibrillin-1 and fibrillin-2 suggesting a role for MAGP-2 in elastic fiber assembly. *J Biol Chem* 277(38):35044–35049.
2. Miyamoto A, Lau R, Hein PW, Shipley JM, Weinmaster G (2006) Microfibrillar proteins MAGP-1 and MAGP-2 induce Notch1 extracellular domain dissociation and receptor activation. *J Biol Chem* 281(15):10089–10097.
3. Mok SC, et al. (2009) A gene signature predictive for outcome in advanced ovarian cancer identifies a survival factor: Microfibril-associated glycoprotein 2. *Cancer Cell* 16(6):521–532.
4. Buchholz M, et al. (2005) Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions. *Oncogene* 24(44):6626–6636.
5. Chagnon MJ, Uetani N, Tremblay ML (2004) Functional significance of the LAR receptor protein tyrosine phosphatase family in development and diseases. *Biochem Cell Biol* 82(6):664–75.
6. Konishi N, et al. (2003) Overexpression of leucocyte common antigen (LAR) P-subunit in thyroid carcinomas. *Br J Cancer* 88(8):1223–1228.
7. Yang T, Zhang JS, Massa SM, Han X, Longo FM (1999) Leukocyte common antigen-related tyrosine phosphatase receptor: increased expression and neuronal-type splicing in breast cancer cells and tissue. *Mol Carcinog* 25(2):139–149.
8. Krupenko SA (2009) FDH: an aldehyde dehydrogenase fusion enzyme in folate metabolism. *Chem Biol Interact* 178(1-3):84–93.
9. Black WJ, et al. (2009) Human aldehyde dehydrogenase genes: Alternatively spliced transcriptional variants and their suggested nomenclature. *Pharmacogenet Genomics* 19(11):893–902.
10. Chen XQ, He JR, Wang HY (2012) Decreased expression of ALDH1L1 is associated with a poor prognosis in hepatocellular carcinoma. *Med Oncol* 29(3):1843–1849.
11. Feng L, Hatten ME, Heintz N (1994) Brain lipid-binding protein (BLBP): A novel signaling system in the developing mammalian CNS. *Neuron* 12(4):895–908.
12. Liang Y, et al. (2005) Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci USA* 102(16):5814–5819.
13. Slipicevic A, et al. (2008) The fatty acid binding protein 7 (FABP7) is involved in proliferation and invasion of melanoma cells. *BMC Cancer* 8:276.
14. Seliger B, et al. (2005) Identification of fatty acid binding proteins as markers associated with the initiation and/or progression of renal cell carcinoma. *Proteomics* 5(10): 2631–2640.
15. Liu RZ, et al. (2012) A fatty acid-binding protein 7/RXRβ pathway enhances survival and proliferation in triple-negative breast cancer. *J Pathol* 228(3):310–321.

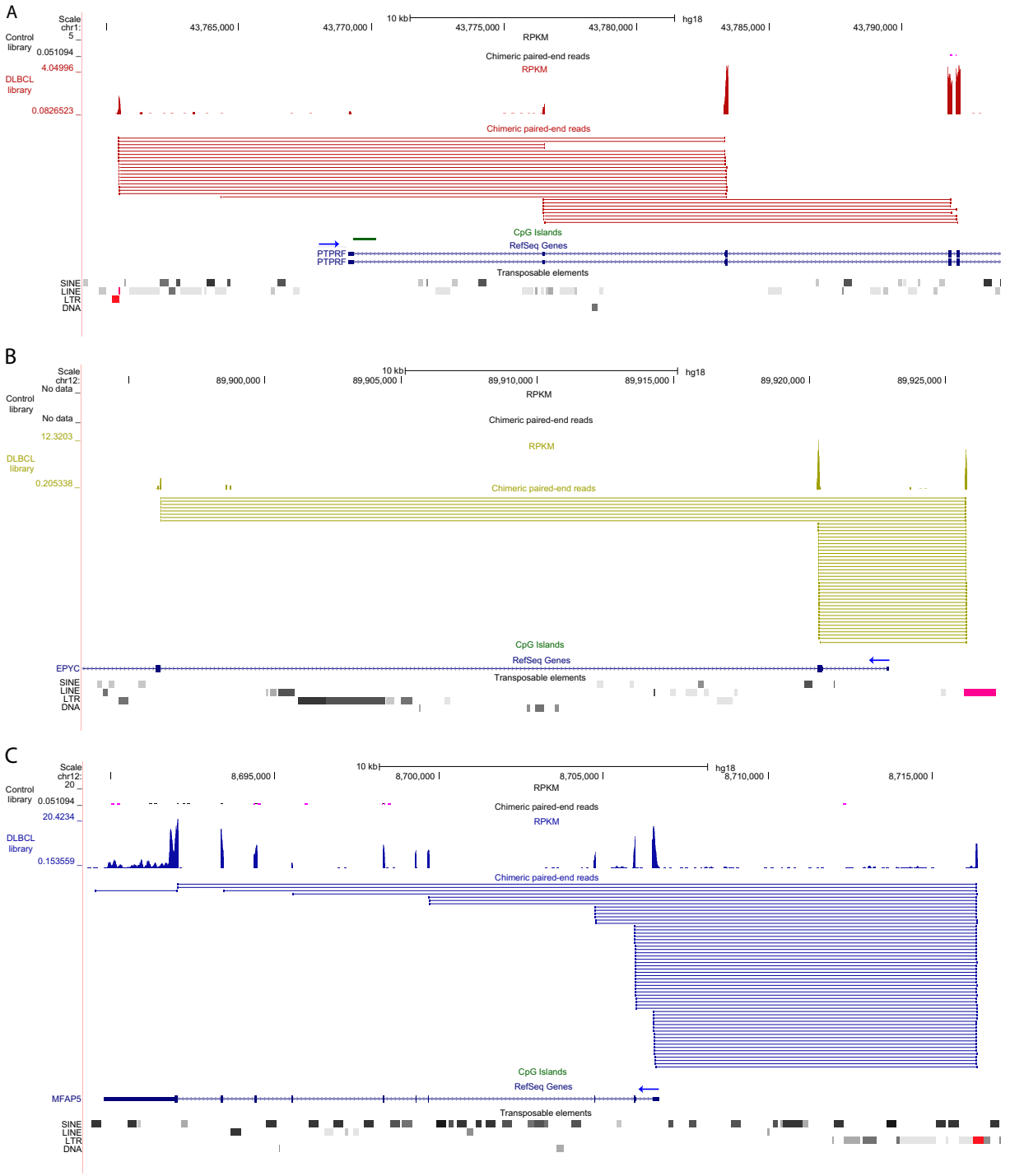

Fig. S1. (Continued)

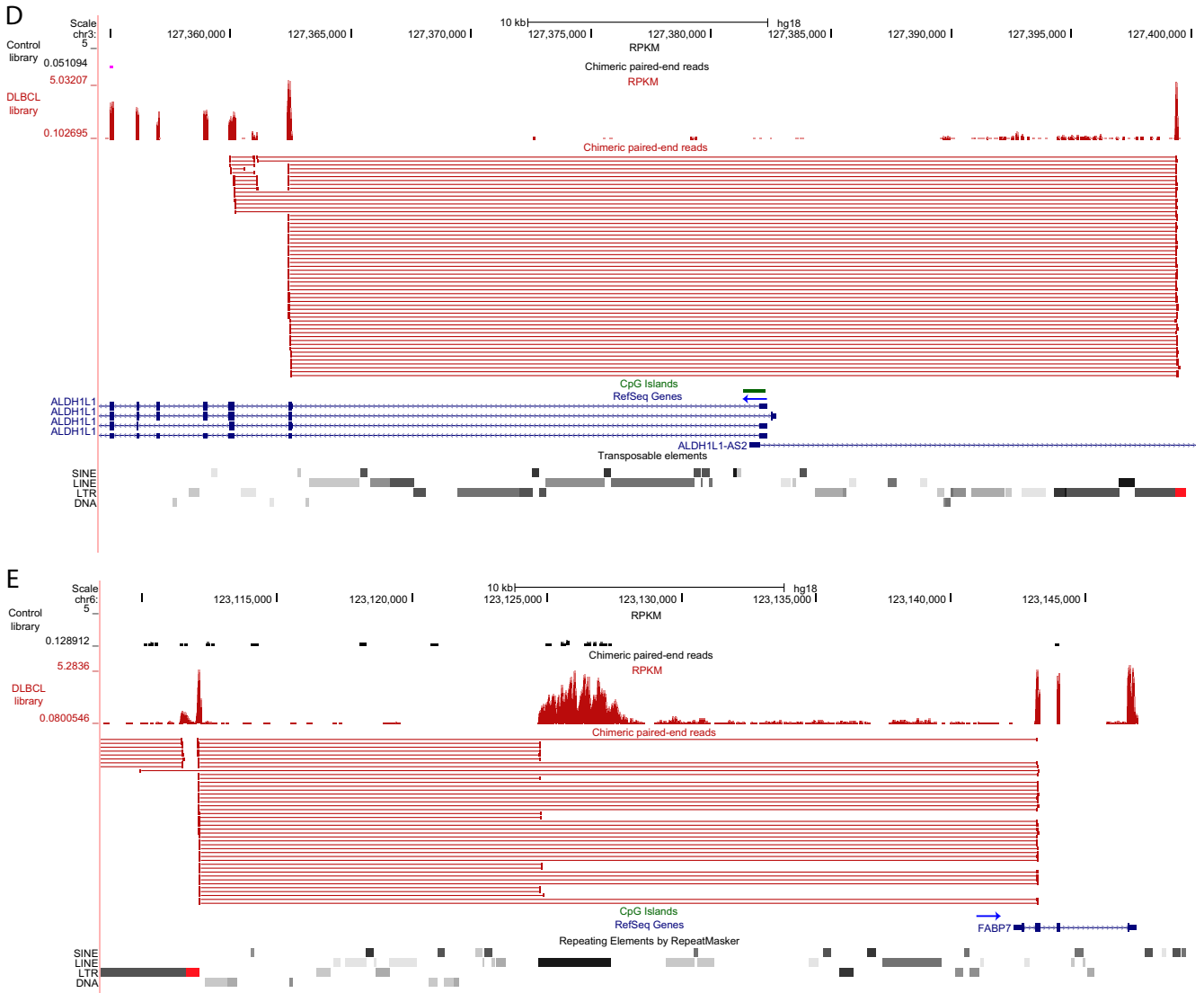

Fig. S1. Human hg18 UCSC genome browser screenshots for the five transposable element (TE)-gene chimeric candidates modified to show only paired-end reads verified to be chimeric (one read located in a TE sequence). RPKM tracks are computed with all paired-end reads including reads mapping to multiple locations but where mates are uniquely mapped. TEs acting as promoters are depicted in red or pink and the sense of gene transcription is showed as a blue arrow on top of each gene: (*A*) *PTPRF*, (*B*) *EPYC*, (*C*) *MFAP5*, (*D*) *ALDH1L1*, and (*E*) *FABP7*. The DLBCL library depicted in the *FABP7* screenshot is different from in Fig. 1*A* to show different libraries for this example.

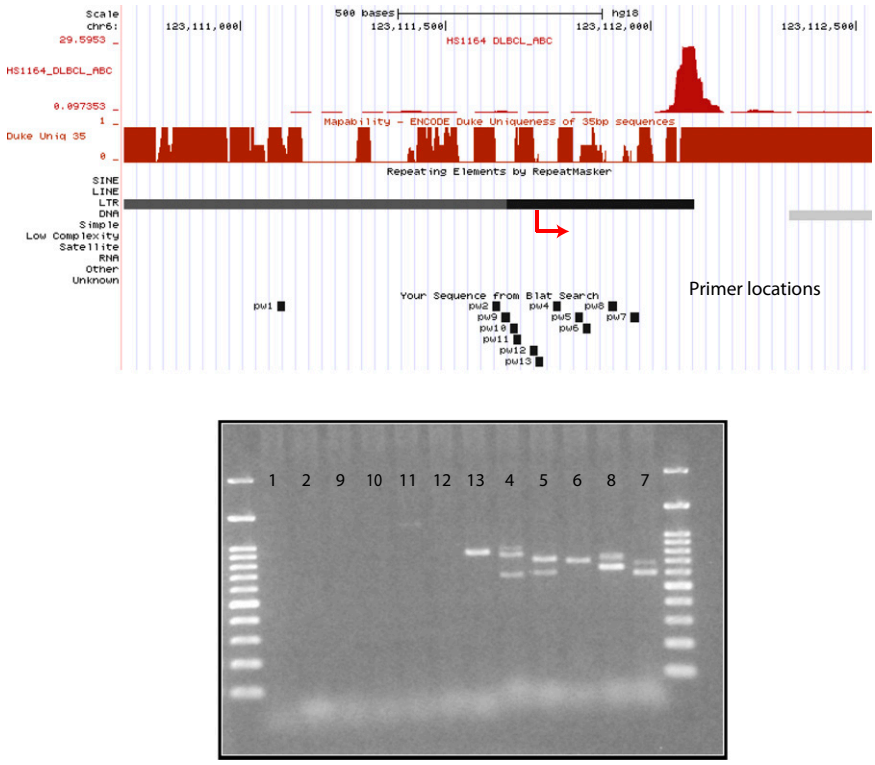

Fig. S2. Primer walking for LTR2-*FABP7*. Hg18 UCSC genome browser screenshot of the LTR2 upstream of *FABP7* putatively acting as a promoter showing the peak of reads for one DLBCL sample. Note that the mappability of the region around the localized transcription start site is low, resulting in a lack of reads mapping closer to it. The RT-PCR product obtained to confirm presence of LTR2-*FABP7* chimera is shown along with all of the primers used for primer walking in the Blat track. On the RT-PCR gel, a 100-bp ladder is shown and numbers correspond to primer names in the genome browser screenshot.

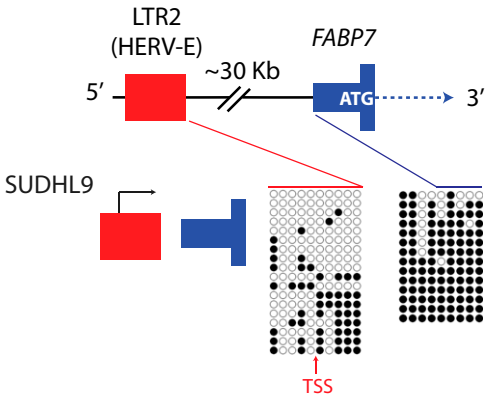

Fig. S3. Bisulfite sequencing of native and TE promoters for *FABP7* in the SUDHL9 DLBCL cell line. Gene scheme is the same as in Fig. 1*B*; filled circles represent methylated CpGs, empty circles represent unmethylated CpGs, and each row represents an independent clone. The localization of CpGs relative to the DNA sequence is shown with colored lines. This cell line was inconsistently positive for the chimeric transcript by RT-PCR but negative for expression from the native promoter, and protein expression was not detectable (Fig. 3*A*).

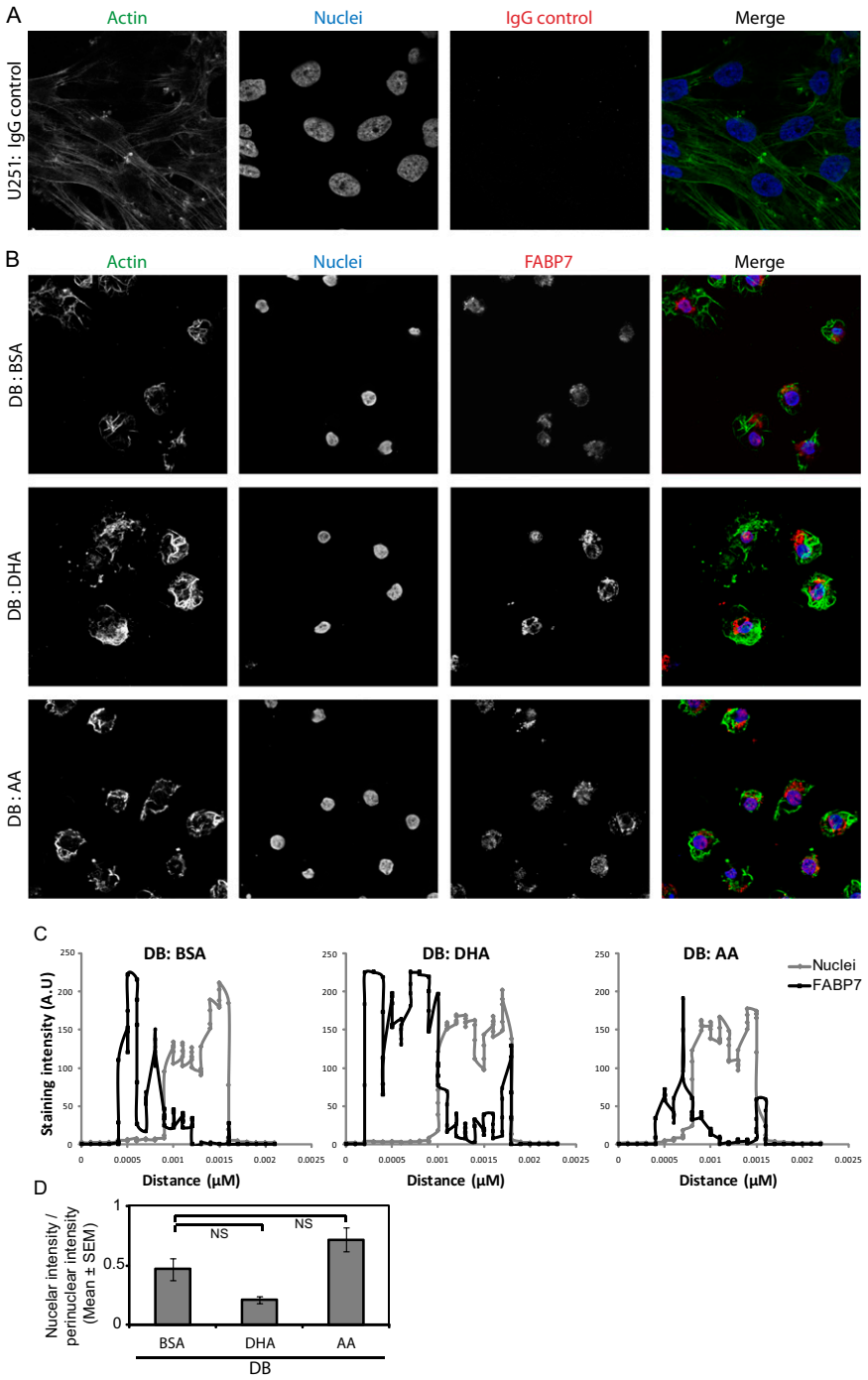

Fig. S4. (*A*) U251 cells or (*B*) DB cells were treated with BSA (vehicle), ω-3-docosahexaenoic acid (DHA), or ω-6-arachidonic acid (AA), fixed, and stained with phalloidin (Actin), FABP7 antibodies, or an IgG control, as specified, and DAPI nuclear stain. FABP7 localization was assessed by confocal microscopy. (*C*) Representative data of FABP7 and DAPI staining intensity across the diameter of a single DB cell. A.U., arbitrary units. (*D*) Mean nuclear FABP7 fluorescence intensity/perinuclear FABP7 fluorescence intensity from at least 30 DB cells is given.

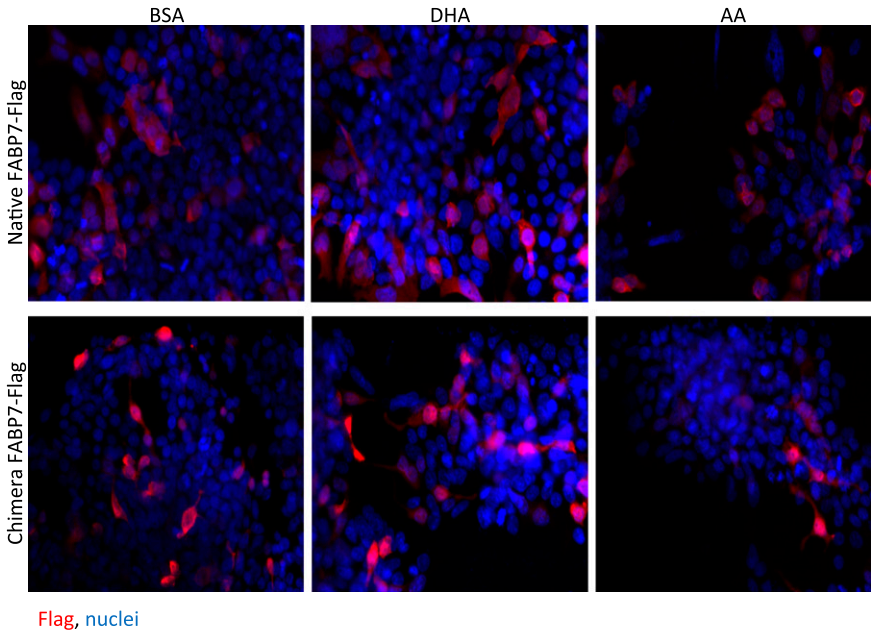

Fig. S5. cDNAs encoding native or chimeric FABP7 were cloned into the pCMV-3Tag (Flag) expression vector (Agilent), transfected into HEK293 cells by calcium phosphate precipitation, and treated with BSA (vehicle), DHA, or AA as previously. Transfected FABP7-Flag localization was assessed by Flag immunofluorescence.

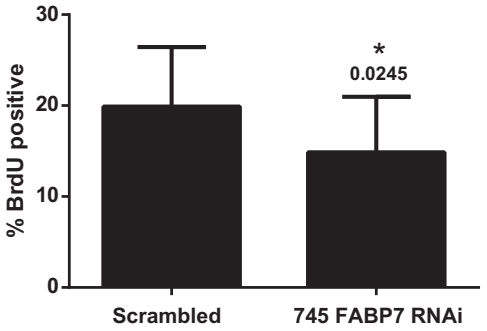

Fig. S6. SUDHL4 were stably transfected with RNAi vectors targeting FABP7 (745) or a scrambled control and subjected to BrdU incorporation assay, assessed by FACS analysis. Representative of two independent experiments.

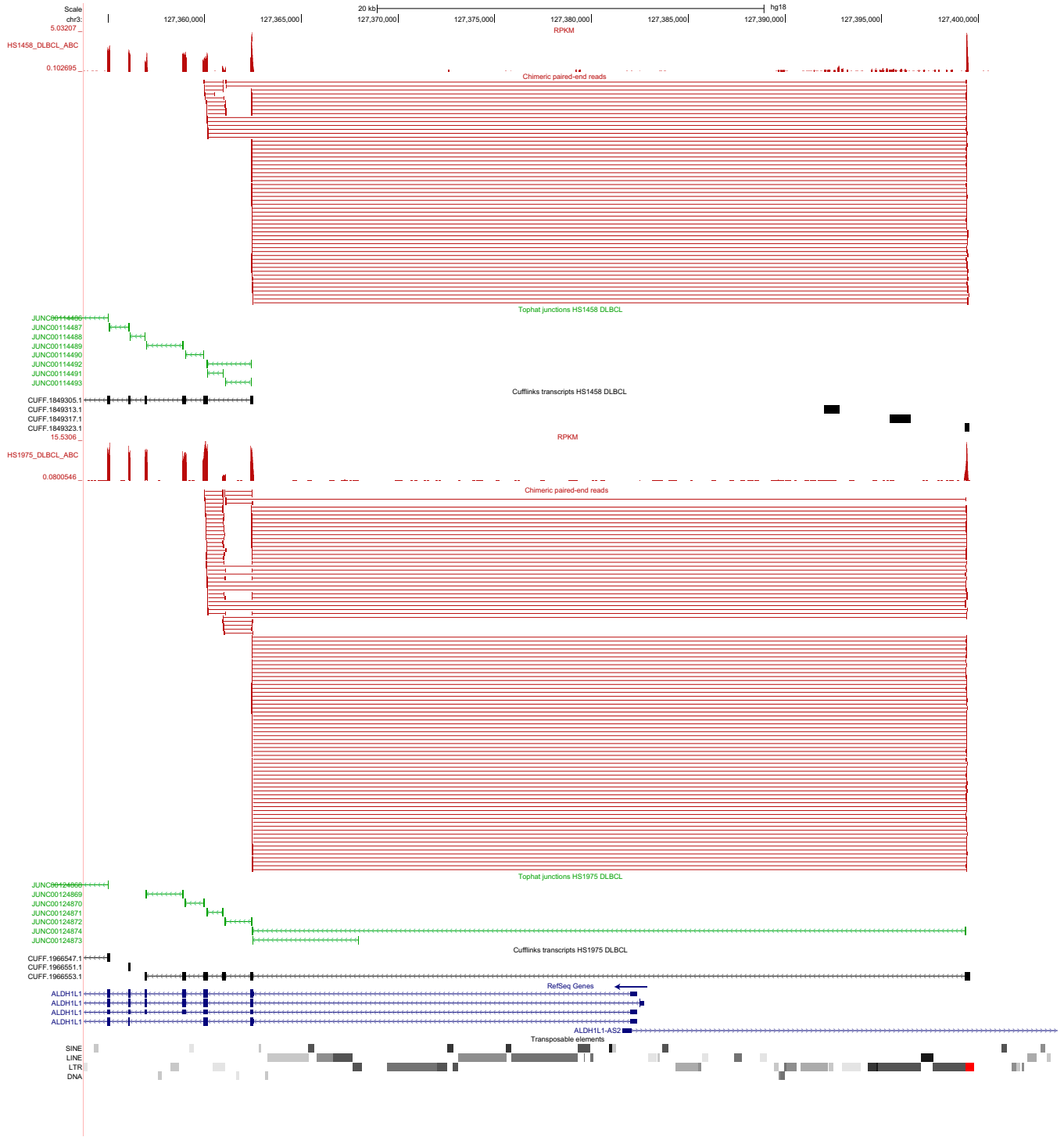

Fig. S7. Example of our paired-end chimeric read analysis compared with ab initio transcript assembly using Tophat/Cufflinks. Human hg18 UCSC genome browser screenshots for the *ALDH1L1* gene region in two DLBCL libraries. For each library, the tracks show reads per kilobase of transcript per million reads mapped (RPKM), chimeric paired-end reads found by our analysis in red as well the result of ab initio transcript assembly using split read junctions in green (Tophat) and assembled Cufflinks transcripts in black. The LTR acting as the promoter is highlighted in red in the transposable element (RepeatMasker) track and the sense of gene transcription is showed as a blue arrow above the Refseq track. The LTR-initiated transcript is not detected by ab initio assembly in the upper library.

Table S1. Primers used in this study: 5′ RACE

ALDH1L1	Primer sequence
RLM 5′RACE inner primer	CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG
ALDH1L1 exon 2 forward	CCTCAGGTGGCAGTAAACTTC

Table S2. Primers used in this study: RT-PCR

Target	Forward	Reverse
GABPDH	CATGAGAAGTATGACAACAGCCTC	GTTGCTGTAGCCAAATTCGTTGTC
FABP7 native	GCTACCTGGAAGCTGACCA	CTGAAGAGCTCTTCCAAGCC
FABP7 chimera	TGGAATGCAACTGAAGGAGA	CTGAAGAGCTCTTCCAAGCC
FABP7 total	TGGATGGAGACAAACTTGTTCA	TATGCCTTCTCATAGTGGCGA
EPYC native	GCTCTGCATCCTCAGTCACT	TTCTTTGGCAGCGGAGGAAT
EPYC chimera	GCCAACCAACCACCACTTTG	TTCTTTGGCAGCGGAGGAAT
EPYC total	ATTCCTCCGCTGCCAAAGAA	GTGGCAGAGGGATGTGGTC
PTPRF native	TGGGTTCTCCTGTAGCTTGG	GCAGAGGAGTGGGAGCAG
PTPRF chimera	GGGAAACCTGACCTAACACAGA	TGCTGGATACTCAGCACCAA
PTPRF total	GAGGACCAGACTGGGCTGT	GAACGCAGCTGCTTGATG
ALDH1L1 native	AGTGAACACACCCACCACTTC	GGTCTGCACCAACCTCAGTT
ALDH1L1 chimera	AGTGAACACACCCACCACTTC	TTGTGAATCGTTCATTGCTC
ALDH1L1 total	TAACTCCAGGCCATCACACA	AGAGGCCATTCACAACTGGA
MFAP5 native	GTGCAATATCAGCCAAA	ATTCCAGCCTCATTG
MFAP5 chimera	ATGCAGCAAGAAACAGCAGC	TAAGGTCTCCCCAGCCATGT
MFAP5 total	AAGGAAGACTTGCCAGGCAA	CGGCCGGTTAAACAATGCAT

Table S3. Primers used in this study: Bisulfite analysis

Gene promoter	Primer sequence	Primer sequence
MFAP5 native, first round	MFAP5 BIS-S1 TTTAGGGTTTTAGGTAATTAAATAATTTAT	MFAP5 BIS-AS1 ACAAAAAAAATCTAAATATATAAATCAAAA
MFAP5 native, second round	MFAP5 BIS-S1 TTTAGGGTTTTAGGTAATTAAATAATTTAT	MFAP5 BIS-AS5 AATCACATCATTACACTCCAACCTA
MFAP5 LTR, first round	MFAP5 BIS-S2 TTTAATTTTTGTTTGTTATTTAGTTTTAAA	MFAP5 BIS-AS3 AAAACAAAACAAAAACAAACCATTC
MFAP5 LTR, second round	MFAP5 BIS-S2 TTTAATTTTTGTTTGTTATTTAGTTTTAAA	MFAP5 BIS-AS4 AAAATCATCTTTCCCATACTATTCTC
FABP7 native, first round	FABGeneBisulF GGGGTTTTTATTTGTGATTAGTTTT	FABGeneBisulR1 TTATTACCTACCTAAAACCTTCATATACTC
FABP7 native, second round	FABGeneBisulF GGGGTTTTTATTTGTGATTAGTTTT	FABGeneBisulR2 CTCATCAAAATTCTAACTATTAATCAACTT
FABP7 LTR, first round	FABLTRBisulF TTGGTTAGAGTTTTTTGTAGGGATG	FABLTRBisulR1 TTCAATTACATTCCACCAAATTTAAA
FABP7 LTR, second round	FABLTRBisulF TTGGTTAGAGTTTTTTGTAGGGATG	FABLTRBisulR2 AAAAAAAACCACCCCTCATATTATC
ALDH1L1 native, first round	ALDH1L1geneF1 GGGTAGGTTAGATTTTTGTGAGT	ALDH1L1geneF2 GAAATTGAGGTTGGTGTAGATT
ALDH1L1 native, second round	ALDH1L1geneF1 GGGTAGGTTAGATTTTTGTGAGT	ALDH1L1geneR AAAAACCAAAAACTATTTTTCTCTC
ALDH1L1 LTR, first round	ALDH1L1LTRF1 TTATTTTGAGAGATTAATGGATATAAATAG	ALDH1L1LTRF2 TTTGATATTATGTTAAAAGAAAAATTTTAG
ALDH1L1 LTR, second round	ALDH1L1LTRF1 TTATTTTGAGAGATTAATGGATATAAATAG	ALDH1L1LTRR ACAAAAACTAAAACAAACAAAAAAC

Table S4. Primers used in this study: Primer walking

Primer	Sequence
Reverse primer	
orfF4Rev	CTGAAGAGCTCTTCCAAGCC
Forward primers	
FABP7 pw-1	TGCCTGACCCATATTTCCTC
FABP7-pw2	AATCACCACGTTTCCTGGTC
FABP7-pw3	GGGTTGTCTGCTTGTGGATT
FABP7 pw4	TGAGCAATTCCTGTCCCTTT
FABP7 pw5	AGGGTCGTGATTGATTGAGC
FABP7 pw6	CAAGCAGGGGGTACATGACT
FABP7 pw7	CCAATTAGGTCAGGGGTTGA
FABP7 pw8	CAGGGATTTTCACAGTGCTTT
FABP7 pw9	GGAACCAAGAAATGTAGCAGGA
FAPB7 pw10	GACAAGCCGCAGACAAAACC
FABP7 pw11	GCAGACAAAACCCCTCAGAC
FABP7 pw12	GTTTATTCCACCGGGAGCAT
FABP7 pw13	GGAGCATCAGCAAGACTCCT

Dataset S1. Bioinformatics pipeline output of the 98 chimeric genes where the gene is silent in normal (RPKM <1) and the following criteria are met in at least two DLBCL libraries: (*i*) presence of at least three chimeric reads and (*ii*) expression in DLBCL libraries (RPKM ≥1)

Dataset S1

All of the different RefSeq symbols with the number of chimeric reads and RPKM for each library where the parameters are met are shown. Each Refseq gene name is shown by alternating colors.