NEOPLASIA
www.neoplasia.com

Volume 14 Number 11 November 2012 pp. 1087–1096 1087

Long-range Transcriptome Sequencing Reveals Cancer Cell Growth Regulatory Chimeric mRNA[1,2]

Roberto Plebani[*,3], Gavin R. Oliver[†,3], Marco Trerotola[*,4], Emanuela Guerra[*], Pamela Cantanelli[*], Luana Apicella[*], Andrew Emerson[‡], Alessandro Albiero[§], Paul D. Harkin[†,¶], Richard D. Kennedy[†,¶] and Saverio Alberti[*,#]

*Unit of Cancer Pathology, Centre of Excellence for Research on Aging, "G. D'Annunzio" University Foundation, Chieti, Italy; [†]Almac Diagnostics, Craigavon, United Kingdom; [‡]CINECA, Bologna, Italy; [§]BMR Genomics, Padova, Italy; [¶]Centre for Cancer Research & Cell Biology, Queen's University, Belfast, United Kingdom; [#]Department of Neuroscience and Imaging, "G. d'Annunzio" University of Chieti-Pescara, Chieti, Italy

Abstract

mRNA chimeras from chromosomal translocations often play a role as transforming oncogenes. However, cancer transcriptomes also contain mRNA chimeras that may play a role in tumor development, which arise as transcriptional or post-transcriptional events. To identify such chimeras, we developed a deterministic screening strategy for long-range sequence analysis. High-throughput, long-read sequencing was then performed on cDNA libraries from major tumor histotypes and corresponding normal tissues. These analyses led to the identification of 378 chimeras, with an unexpectedly high frequency of expression ($\approx 2 \times 10^{-5}$ of all mRNA). Functional assays in breast and ovarian cancer cell lines showed that a large fraction of mRNA chimeras regulates cell replication. Strikingly, chimeras were shown to include both positive and negative regulators of cell growth, which functioned as such in a cell-type–specific manner. Replication-controlling chimeras were found to be expressed by most cancers from breast, ovary, colon, uterus, kidney, lung, and stomach, suggesting a widespread role in tumor development.

Neoplasia (2012) 14, 1087–1096

Introduction

Several chimeric transcripts have been discovered in human solid tumors, which derive from chromosomal translocations. These often encode structurally and functionally altered signaling molecules or transcription factors [1] or may also function as non-coding RNA [2]. More than half of prostate cancers harbor fusion sequences, mostly *TMPRSS-ERG* [3]. The *SLC45A3-ELK4* (ETS family) fusion transcript can be generated both by chromosomal rearrangement and by trans-splicing, and it was found to be expressed in both normal prostate tissue and in prostate cancer. High levels of *SLC45A3-ELK4* mRNA are restricted to a subset of prostate cancer samples [4]. A small inversion within chromosome 2p leads to the formation of a fusion gene comprising *EML4* and *ALK* in non–small cell lung cancer [5]. The fusion of *MAML2* with *CRTC1* or *CRTC3* has a role in the development of mucoepidermoid carcinomas [6]. Rearrangements of *RAF* pathway members occur in prostate and gastric cancers [7], and a paracentric inversion of chromosome 7q results in an in-frame fusion

Abbreviations: FP, fusion point; NGS, next-generation sequencing; SD, standard deviation

Address all correspondence to: Prof. Saverio Alberti, MD, PhD, Unit of Cancer Pathology, Ce.S.I., University "G. d'Annunzio", Via Colle dell' Ara, 66100 Chieti Scalo (Chieti), Italy. E-mail: s.alberti@unich.it

[1]This work was supported by Fondazione Cassa di Risparmio della Provincia di Chieti, Italian Ministry of Health (RicOncol grant RF-EMR-2006-361866), Fondazione Compagnia di San Paolo (grant 2489IT), Ministero dello Sviluppo–Made in Italy (contract N° MI01_00424), and the Italian Foundation for Cancer Research (fellowship to M.T.).

[2]This article refers to supplementary materials, which are designated by Tables S1 to S11 and Figures S1 to S4 and are available online at www.neoplasia.com. The FusionMiner software is freely available at FusionMiner.sourceforge.net.

[3]These authors contributed equally to this work.

[4]Current address: Kimmel Cancer Center, Department of Cancer Biology, Thomas Jefferson University, Philadelphia, PA 19107.

Received 16 August 2012; Revised 16 August 2012; Accepted 30 September 2012

 1522-8002/12/$25.00
DOI 10.1593/neo.121342

between exons 1 and 8 of the *AKAP9* gene and between exons 9 and 18 of *BRAF* in radiation-induced papillary carcinomas [8]. Other thyroid carcinoma–specific events include fusion of the *RET* oncogene to various partners [9]. Further oncogenic fusions have been detected in other solid tumors [10,11].

Cancer transcriptomes also contain mRNA chimeras that arise as transcriptional (long intergenic transcription) or post-transcriptional (trans-splicing [12]) events that may play a role in tumor development. Previous findings showed that oncogenic transcripts can indeed be generated post-transcriptionally [13–15]. The fusion of *CYCLIN D1* mRNA to *TROP2* transcripts generates oncogenic *CYCLIN D1-TROP2* chimeras, whose tumor-promoting function is induced with a dramatically increased mRNA stability [13]. The oncogenic *JAZF1-JJAZ1* chimeric mRNA can be originated by trans-splicing as well as by a chromosomal translocation [14]. Similarly, the *SLC45A3-ELK4* chimeric transcript can be generated in the absence of chromosomal rearrangements [4,16]. Intergenic splicing generates a ubiquitous chimeric mRNA between the *P2Y11* and *SSF1* transcripts [17]. The generation of these chimeras appears as a regulated event [13,14] and was shown to also occur in normal tissues [4,13,14,17–20]. Several of these chimeric transcripts have been used as diagnostic or prognostic [21] markers and as targets for anti-neoplastic therapy [10,13,22,23].

Screening strategies were previously developed for *in silico* identification of mRNA chimeras in cancer cells [24]. Next-generation sequencing (NGS) approaches now provide much larger sequence information for chimera discovery [7,19,20,25–27]. However, most second-generation NGS approaches generate highly multiplexed, short-tag sequence reads, which are then condensed in strings of base-call probabilities, through a probabilistic fitting of massively parallel data sets. This makes contig assemblies and target alignments correspondingly more difficult [19,25,27,28]. Alignments to complex genomes are even more hampered, because of higher sequence complexity [29] and homology within closely related gene families and pseudogenes.

These problems have led to significant efforts for achieving longer sequence reads and higher sequencing accuracy. In 2005, 454 launched the first NGS apparatus, which was able to generate 100-bp reads. Sequence reads extended to 200 bp in 2007 [30] and are close to 900 bp at present [31]. SOLID sequencing generated 35-bp reads in 2007 [30], and this extended to 75 bp in 2011 [32]. Illumina generated 36-bp sequence reads in 2006 to 2008 [30]. These extended to 100 bp in 2010 [31] and to 300 bp in 2012 (www.illumina.com). Ion Torrent introduced its first sequencer at the end of 2010, and this was capable of 100-bp-long reads. As of 2012, reads of 525-bp average length have been obtained (www.iontorrent.com/lib/images/PDFs/pe_appnote_v12b.pdf). Pacific Biosciences (www.pacificbiosciences.com) succeeded in obtaining even longer reads, which currently are up to 1500 bp.

To take advantage of these technical advances, we have developed an analytical strategy for high-accuracy identification of mRNA chimeras in long-read DNA sequence data sets (Figure 1). This strategy was shown to work efficiently for chimera recognition (Tables S1–S7 and Figure S1). High-throughput, long-read sequencing was then performed on cDNA libraries from major tumor histotypes and corresponding normal tissues. This led to the identification of 378 chimeras, from both normal and transformed cells, indicating an unexpectedly high frequency of expression ($\approx 2 \times 10^{-5}$ of all mRNA). Functional assays in breast and ovarian cancer cell lines showed that a large fraction of mRNA chimeras regulate cell replication. Strikingly, chimeras were shown to include both positive and negative regulators of cell growth, which functioned as such in a cell-type–specific manner. Replication-controlling chimeras were found to be expressed by most cancers from breast, ovary, colon, uterus, kidney, lung, and stomach, suggesting selective pressure for a role in tumor development.

Materials and Methods

Cells

Human MCF-7, MCF-7/Almac, HBL-100, SK-BR-3, MDA-MB-231, MDA-MB-361, MDA-MB-415, MDA-MB-453, MDA-MB-468, HS578, and ZR751 breast cell lines and SKOV-3, IGROV-1, OVCAR-3, and OVCA-432 ovarian cancer cell lines were grown in RPMI 1640 medium supplemented with 10% FBS, 100 IU/ml penicillin, and 100 μg/ml streptomycin (Euroclone, Milan, Italy). All cell lines were obtained from ATCC (LGC Standards, Teddington Middlesex, United Kingdom) where they were authenticated by standardized procedures (www.atcc.org).

Cell Growth Assays

MCF-7, HBL-100, SK-OV-3, IGROV-1, and OVCAR-3 cells were seeded at 1×10^3 to 10×10^3 cells/well in 96-well plates (five replicates per data point). Cell numbers were quantified by staining with crystal violet [33]. Standard growth curves for each cell line were generated by seeding two-fold serial dilutions of defined cell numbers. Crystal violet standard curves showed good linear responses ($R^2 > 0.998$, in all cases) (Figure S2). To support the crystal violet readings, quantification was also performed by image analysis (ImageJ). Digital pictures were taken from 96-well plates after fixation. Picture noise was removed with GIMP software, after random sampling of cell-free pixels. ImageJ analysis was then performed by quantifying black areas in each culture well after image conversion to a gray scale (manuscript in preparation).

DNA Transfection

Cells were transfected with DNA in Lipofectamine 2000 (HBL-100, SKOV-3, IGROV-1, and OVCAR-3 cells) or LTX (Invitrogen, San Diego, CA), which was found to be optimal for MCF-7 cells (Figure 5*C*) [34], following the manufacturer's instructions. pEYFP transfection was used to quantify transfection efficiency [35] (EYFP expression, as measured by flow cytometry).

Flow Cytometry Immunofluorescence

Flow cytometry analysis was performed as described previously [36,37], on fluorescence-activated cell sorters (FACSCalibur, Becton-Dickinson, Sunnyvale, CA). To improve the detection of EYFP transfectants, we performed subtraction of cell autofluorescence and displacement of true transfectants in the red channel as described [35,38].

Human Samples for Tumor Transcriptome Sequencing

Non–small cell lung cancer. Non–small cell lung cancer libraries were generated from a set of frozen tissue samples, comprising 65 tumor samples (30 adenocarcinomas, 20 squamous cell carcinomas, and 15 other morphologies) from the Roy Castle Lung Cancer Research Institute (University of Liverpool) and Queens University Belfast. To maximize chances of mRNA chimera discovery, we proceeded to generate libraries from both tumor and normal tissues. Normal lung RNA was obtained from multiple commercial suppliers (Clontech, Palo Alto, CA; Ambion, Austin, TX; BioCat, Heidelberg, Germany; Stratagene, La Jolla, CA; Cybridi, Rockville, MD; and OriGene, Rockville, MD), overall from 16 donors of different ethnicity.

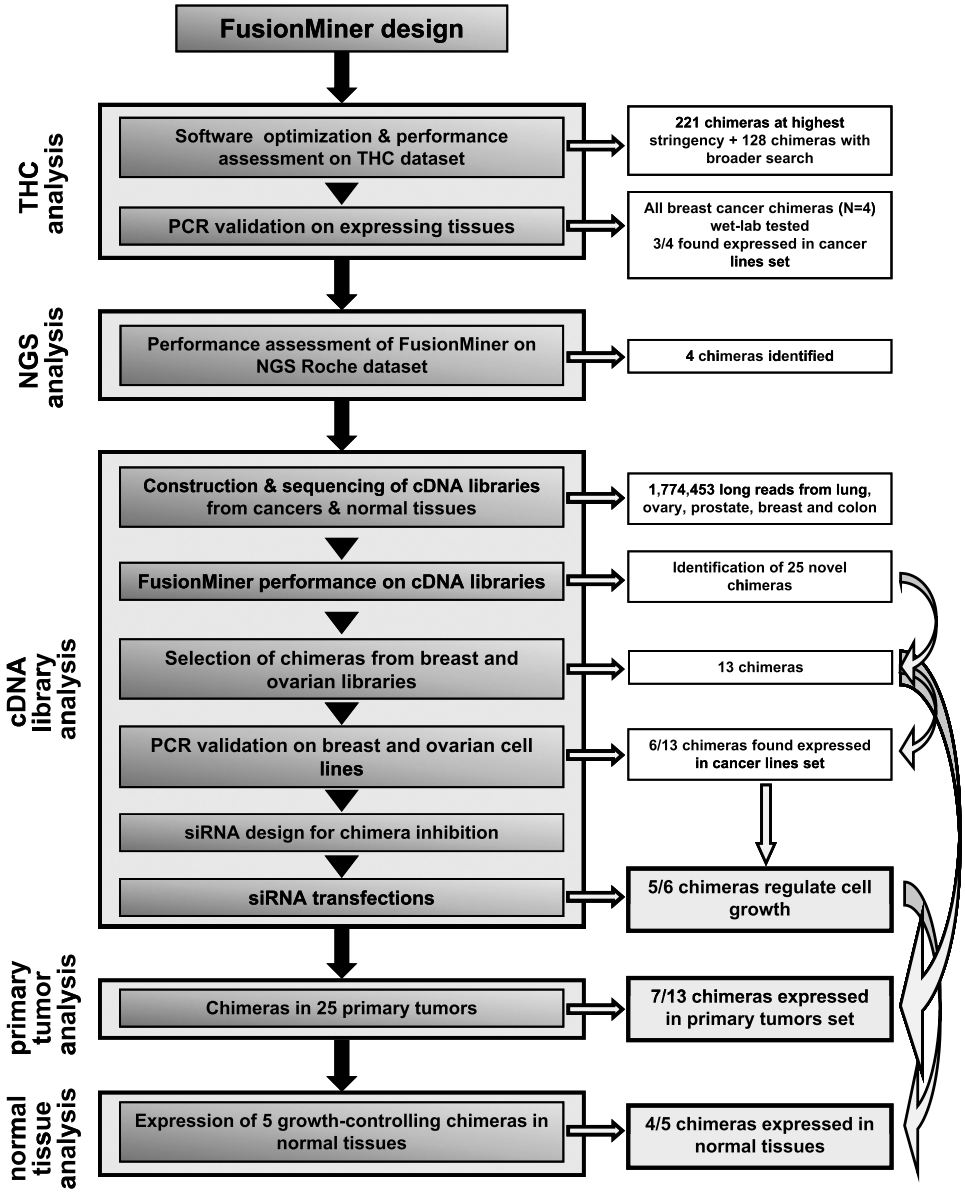

Figure 1. Flow diagram of chimera identification and validation steps.

Ovarian cancer. Ovarian library A comprised 64 ovarian tumors (31 serous, 14 endometroid, 6 mucinous, 5 clear cell, and 8 undifferentiated cancers; 52 were stage III/IV and 12 were stage I/II). For ovarian library B, RNA from normal ovarian tissue was obtained from commercial suppliers (Ambion and AMS Biotechnology [Bioggio, Switzerland]). The library was generated with equal quantities of RNA from different ethnicities (Asian, Caucasian, and African-American), with 23 donors overall. For ovarian library C, ovarian tumor total RNA was obtained from various commercial suppliers (Ambion, Clontech, Cytomyx [Lexington, MA], Biocat, and Asterand [Detroit, MI]). The library was composed of equal quantities of RNA of different ethnicity (Asian, African-American, and Caucasian), with 37 donors overall.

Prostate cancer. The prostate cancer library was constructed from 30 tumors (74% Caucasian and 26% African-American), 8 normal prostate RNA supplied by Clontech, AMS Biotechnology, and Cybridi, and 56 normal tissues adjacent to tumors obtained from St. Vincents Hospital (Dublin, Ireland).

Breast cancer. The breast cancer library was composed of 90 tumors and 18 normal samples [39–41].

Colorectal cancer. The colorectal library comprised 40 tumor samples and 40 normal tissues.

Tumor Validation Sample Set

cDNA was synthesized from 25 human primary tumors (10 breast, 6 colon, 3 stomach, 2 ovary, 2 kidney, and 2 uterus), which were independent from those used to construct the cDNA libraries. These 25 samples were used as a test set to validate chimera expression by both conventional polymerase chain reaction (PCR)/sequencing and real-time reverse transcription (RT)–PCR.

Normal Tissues

Normal breast, colon, uterus, prostate, placenta, lung, kidney, pancreas, and stomach RNA were obtained from Clontech.

cDNA Library Construction

All of the frozen tumor tissues were homogenized in RNA STAT-60 (Tel-Test, Friendswood, TX), and the RNA was extracted according to the manufacturer's instructions. Equal amounts of good quality total RNA were pooled, and the mRNA was isolated using μMACS mRNA isolation kits (Miltenyi Biotec, Bergisch Gladbach, Germany), as described by the manufacturer. Lung cDNA libraries were constructed from 3 μg of mRNA using the CloneMiner cDNA library construction kit (Invitrogen), according to the manufacturer's instructions. cDNA were inserted in the pDONR 222 vector from Invitrogen. Titer and average insert size in each cDNA library were determined according to the manufacturer's instructions. Plasmid preparations of individual clones were carried out using a modified Montáge alkaline lysis method (Millipore, Billerica, MA) that incorporates MultiScreen Plasmid 384 Miniprep clearing plates for centrifugal lysate clearing.

Sequencing of cDNA Libraries

Colony sequencing automation was implemented (QPix colony picker Biomek liquid handlers). Cycle sequencing reactions were performed in 10-μl volumes using a 1/16 dilution of Big Dye Terminator v3.1 ready reaction mix in Big Dye sequencing buffer (Applied Biosystems, Foster City, CA), 5 μM M13 primer, and 100 ng of template DNA. Cycle sequencing was performed for 40 cycles at 95°C for 10 seconds, 50°C for 5 seconds, and 60°C for 2.5 minutes. Excess dye terminators were removed using CleanSEQ (Agencourt Biosciences Corporation, Beverly, MA). Sequencing plates were analyzed on Applied Biosystems 3730/3730 × l DNA Analyzers using Applied Biosystems Sequence Analysis software. M13 forward primers were used for 5′ end sequencing of the colorectal and breast libraries; M13 reverse primers were used for 3′ end sequencing of the normal lung and prostate libraries; both M13 forward and reverse primers were used for 5′ and 3′ end sequencing of the lung tumor and ovarian cancer libraries.

Plasmids

The pEYFP expression vector (Clontech) was used to express YFP. The pSUPER vector [42] was used for RNA interference.

Small Inhibitory RNA (siRNA)

siRNA design followed four complementary strategies, i.e., Tuschl criteria (position in the mRNA, guanine-cytosine [GC] content, base composition, and flanking sequences) [43], Invitrogen algorithms (rnaidesigner.invitrogen.com/rnaiexpress/; sequence composition, nucleotide content, thermodynamic properties, and experimental validation), Whitehead Institute screening procedures (jura.wi.mit.edu/bioc/siRNAext/; Tuschl criteria, predictions of binding energies and BLAST filtering of cross-hybridizing sequences) [44], and Sonnhammer searches (www.sirnawizard.com/design_advanced.php; data mining on validated siRNA databanks, using motif rules and energy parameters) [45].

Annealed siRNA oligos were subcloned into the pSUPER vector. siRNA expression constructs were transiently transfected in MCF-7 and HBL-100 breast cancer cells and in SK-OV-3, IGROV-1, and OVCAR-3 ovarian cancer cells. siRNA-targeted transcript levels were quantified by real-time PCR. Negative-control siRNA directed toward irrelevant targets were used; these were chosen after extensive testing for lack of off-target influence on cell growth.

Quantitative RT-PCR

Hybrid sequences in cancer cell lines and tumor samples were amplified by quantitative RT-PCR. One microgram of total RNA was reverse transcribed with the M-MLV Reverse Transcriptase (Promega, Madison, WI) according to standard protocols. cDNA was quantified by ethidium bromide fluorescence in solution [46]. Quantitative RT-PCR was performed with an ABI-PRISM 7900HT Sequence Detection System (PE Applied Biosystems, Foster City, CA), using Sybr Green as the probe (Applied Biosystems). Samples were assayed as replicates (two or three independent samples), and the $1.83^{-\Delta\Delta CT}$ method was used to calculate the relative changes in gene expression [13]. The glyceraldehyde 3-phosphate dehydrogenase (GAPDH) housekeeping gene was used as an internal control. For setup curves, ΔC_T (C_T, target gene – C_T, GAPDH) was calculated for each cDNA dilution. The data were fit using least-squares linear regression analysis. As amplification efficiency was linear over the range of RNA amounts used, amplification curves were used to calculate crossover point values for siRNA-treated samples. To check for the correctness of amplified bands, amplification products were run on 3% agarose gels. Amplified products were purified and extensively sequenced (BMR Genomics, Padova, Italy). Quantitative RT-PCR was also performed with PrimeTime IDT (Integrated DNA Technologies, Bologna, Italy; www.idtdna.com) to reliably detect with higher sensitivity the interchromosomal CHD2-CHMP1A fusion in normal tissues.

Diagnostic PCR

Interchromosomal *CHD2-CHMP1A* and *ADK-DHX8* and intrachromosomal *PRKAA1-TTC33*, *SAMM50-PARVB* and *P2RX5-TAX1BP3*, *URB1-C21orf45*, *CTBS-GNG5*, *THC2538403 ZNF498-CUX1*, *THC2523555 C9orf47-S1PR3*, and *THC2668182 KLH22-SCARF* were amplified in 10 breast and 4 ovarian cancer cell lines and in 25 tumor samples by nested PCR. Chimeric mRNA were amplified by 35 amplification cycles (30 seconds at 94°C for denaturation, 30 seconds at 60°C for annealing, and 30 seconds at 72°C for extension). Hot Master Taq-polymerase 0.7 units (Eppendorf) and 12.8 pmol of forward and reverse primers were used for the amplification reaction. All of the amplified products were purified and sequenced (BMR Genomics).

Statistical Analysis

Two-way analysis of variance and *post-hoc* Bonferroni *t* tests were used for growth curve comparisons. Data were analyzed using Sigma Stat (SPSS Science Software UK Ltd, Birmingham, United Kingdom) and GraphPad Prism (GraphPad Software Inc, La Jolla, CA).

Results

Chimeric mRNA Detection Procedure

A procedure (FusionMiner) was designed to process BLAST analyses of query sequences against genomic databanks, through sequential stages of analysis and exclusion and pass-or-fail tests, as described in the Supplemental Online Material (Figure 1 and Tables S1–S7). FusionMiner performance was assessed by screening the Dana Farber Cancer Institute Gene Index Project tentative human consensus (THC) collection (Figure S1) and long-sequence-read 454 Titanium data sets (Supplemental Online Material). Samples of the identified chimeras were then validated by diagnostic PCR and by real-time quantitative PCR analysis of cancer cell lines.

Transcriptome Sequencing for Growth Regulatory Chimera Discovery

To discover growth regulatory chimeras, we then performed a large-scale sequencing and analysis of tumor and normal tissue transcriptomes. To maximize chances of discovery of growth regulatory chimeras, both major tumor histotypes, i.e., non–small cell lung, breast, prostate, ovary, and colorectal cancers, and the corresponding normal tissues were analyzed. Long-sequence-read (900 bp on average) cDNA library data sets were obtained: 481,765 from ovary, 485,049 from prostate, 157,259 from breast, 46,445 from colon, and 603,935 from lung.

These sequences were run through FusionMiner. Twenty-five mRNA chimeras were identified (15 intrachromosomal and 10 interchromosomal; Table S8, Supplemental Sequence Data). All sequences were shown to possess the structural characteristics of *bona fide* chimeric mRNA [24] (Supplemental Sequence Data). Breast and ovarian chimeras were validated by RT-PCR and functional assays (see below).

These findings led to estimate absolute chimera frequencies as 1.4×10^{-5} of all mRNA. This was in remarkable agreement with NGS sequencing data ($\approx 2 \times 10^{-5}$) (Supplemental Online Material), indicating an unexpectedly high frequency of expression of chimeric mRNA.

Chimeric Transcript Expression in Cancer Cells

Expression of the nine chimeras from the breast library and of the four chimeras from the ovarian library was analyzed in breast (MCF-7, HBL-100, SK-BR-3, MDA-MB-231, MDA-MB-361, MDA-MB-415, MDA-MB-453, MDA-MB-468, HS578, and ZR751) and ovarian (SKOV-3, IGROV-1, OVCAR-3, and OVCA-432) cancer cell lines (Figure 2 and Table S8). Six of the nine chimeras were successfully amplified by RT-PCR (Figure 2*A* and Table S9). Amplification from breast cancer cells was obtained for *PRKAA-TTC33* (10/10 lines), *SAMM50-PARVB* (5/10 lines), *P2RX5-TAX1BP3* (3/10 lines), and *CHD2-CHMP1A* (9/10 lines) (Figure 2*A*, *left*). All individual amplicons were sequence verified (Figure 2*C*). Three of these chimeras were also detected in ovarian cancer cells: *PRKAA-TTC33* (4/4 lines); *SAMM50-PARVB* (3/4 lines), and *CHD2-CHMP1A* (4/4 lines) (Figure S3).

The *URB1-C21ORF45* and *CTBS-GNG5* chimeras from the ovarian library were identified in all four ovarian cancer cell lines (Figure 2*A*, *right*). They were also detected in all breast cancer lines. Notably, different cancer cells expressed different steady-state levels of the chimeric mRNA, e.g., *CTBS-GNGS* was approximately 20 times

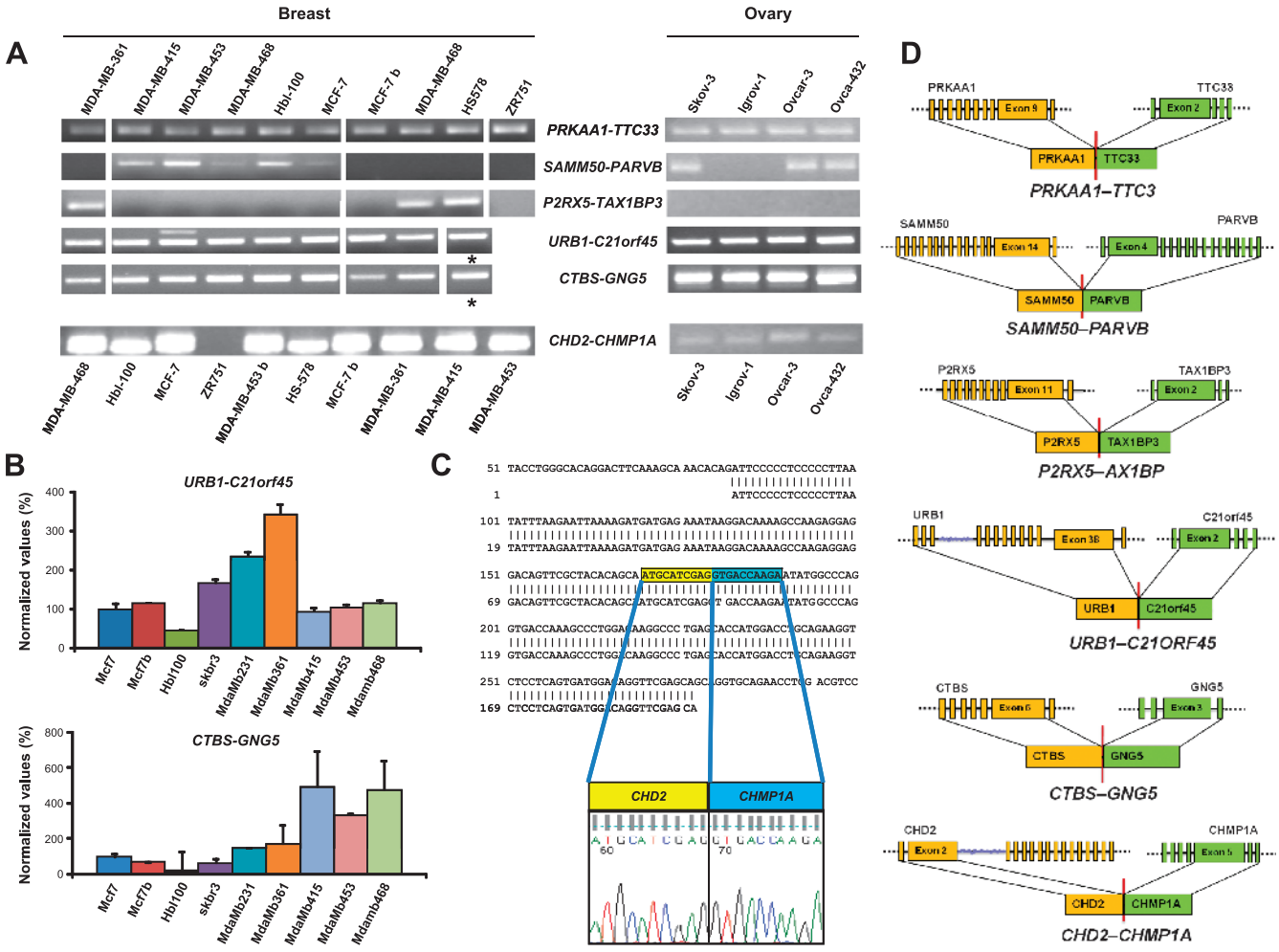

Figure 2. Chimeric mRNA expression in cancer cell lines. (A) Expression of chimeras discovered from tumor and normal tissue library sequencing; agarose electrophoresis of nested or real-time PCR products. Breast and ovarian cancer cell lines are indicated; *SK-BR-3. (B) URB1-C21ORF45 (top) and CTBS-GNG5 (bottom) expression in breast cancer cell lines by quantitative RT-PCR; results are expressed as percent values (MCF-7 = 100); three replica samples were analyzed per data point. Bars, SD. (C) *CHD2-CHMP1A*. Sequence of the PCR amplicon *versus* that of the chimera isolated from the breast library. (D) Structure of validated chimeric mRNA; 5′ partners (orange) and 3′ partners (green) are shown; exon junctions are indicated.

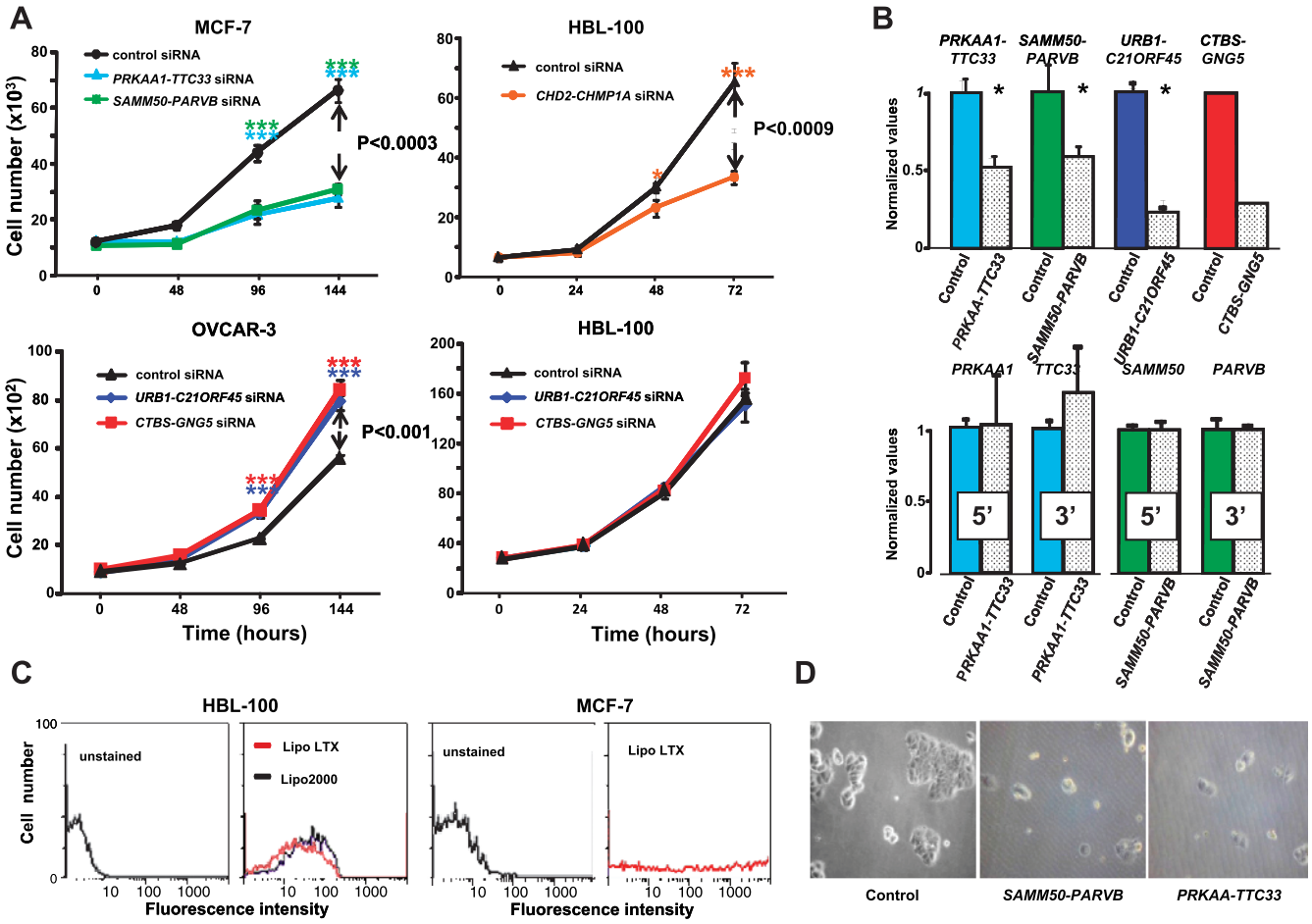

Figure 3. Cell growth modulation by chimeras. (A) Control (Table S10) and chimera-targeting siRNA are color coded. (Top, left) MCF-7; *PRKAA-TTC33* (cyan), *SAMM50-PARVB* (green), control (black). (Top, right) HBL-100; *CHD2-CHMP1A* (orange), control (black). (Bottom, left) OVCAR-3; *URB1-C21ORF45* (blue), *CTBS-GNG5* (red). (Bottom, right) HBL-100 treated with siRNA for chimeras from ovarian libraries; *URB1-C21ORF45* (blue), *CTBS-GNG5* (red). Bars, SD. Brackets, *P* value of two-way analysis of variance; Bonferroni *t* test significance: $*P \leq .05$; $*** P \leq .001$. (B) Real-time PCR of siRNA-transfected cells. (Top) Chimeric RNA (left to right: *PRKAA-TTC33* and *SAMM50-PARVB* in MCF-7; *URB1-C21ORF45* and *CTBS-GNG5* in OVCAR-3). (Bottom) Single-partner RNA expression after the indicated siRNA treatment (left to right: *PRKAA-TTC33* and *SAMM50-PARVB* in MCF-7). (C) Flow cytometry analysis of transfected HBL-100 and MCF7 cells; YFP was used as a transfection efficiency benchmark. (Left) HBL-100; LTX (red) or Lipo-2000 (blue). (Right) MCF-7; LTX transfection. (D) MCF-7 cell growth blockade after *PRKAA-TTC33*–targeted or *SAMM50-PARVB*–targeted siRNA treatment (day 6 after transfection).

less expressed in HBL-100 cells, as compared with MDA-MB-415 cells (Figure 2*B*).

Overall, 75% of the THC chimeras and 54% of the chimeras from breast and ovary libraries (Tumor Transcriptome Sequencing Project) were detected in breast and ovarian cancer cell lines/primary tumors.

Fusion Proteins Encoded by the Growth Regulatory Chimeras

CHD2-CHMP1A. *CHD2* encodes the chromodomain helicase DNA-binding protein 2; *CHMP1A* encodes the chromatin-modifying protein 1A. Of interest, both of these chimera partners encode proteins with regulatory roles on chromatin/DNA structure. However, only the first 20 amino acids of helicase DNA-binding protein 2 are retained in the fusion-protein product (Table S11). This contains a casein kinase II phosphorylation site (prosite.expasy.org/). One out-of-frame C-terminal amino acid is provided by the chromatin-modifying protein 1A sequence (Table S11) and generates a hybrid N-glycosylation site, although it is not clear if this is processed *in vivo*.

CTBS-GNG5. *CTBS* encodes chitobiase; *GNG5* encodes the di-*N*-acetyl–binding and guanine-nucleotide–binding proteins. Chitobiase is a lysosomal glycosidase that is involved in degradation of asparagine-linked oligosaccharides on glycoproteins. It is also involved in the hydrolysis of *N*-acetyl-β-D-glucosamine. *GNG5* encodes the γ chain of trimeric G proteins. A fusion mRNA between chitobiase and guanine-nucleotide–binding protein was also identified by Akiva et al. [47] and by Nacu et al. [26]. The *CTBS-GNG5* is an "in-frame" fusion that preserves the first 319 amino acids from the N-terminal partner and the last 41 amino acids from the C-terminal partner (Table S11). *CTBS* provides an apparently functional chitinase catalytic domain, with a formal glycosylation site at S300. Most of Gγ5 is retained in the fusion (Supplemental Figure S4), which raises the possibility that the fusion protein can bind its Gβ partner, whether at the cell membrane or in the cytoplasm.

PRKAA1-TTC33. *PRKAA1* encodes a 5′-AMP–activated protein kinase catalytic subunit α-1; *TTC33* encodes tetratricopeptide repeat domain 33. *PRKAA1* is a Ser/Thr protein kinase that protects cells from

stress-dependent ATP depletion by switching off ATP-consuming biosynthetic pathways; *PRKAA1* also regulates fatty acid and cholesterol synthesis. The N-terminal segment retained in the chimera contains most of the protein kinase A1 protein (478/559 amino acids; Table S11). This retains the full catalytic domain (46–279). However, it loses 12 phosphorylation sites (T488, T490, T522, S496, S502, S506, S508, S516, S520, S523, S524, and S527), which suggests the loss of at least some of its physiological regulation. The fusion protein contains 32 C-terminal amino acids from the 3′ partner mRNA (*TTC33*), which do not correspond to its canonical reading frame (out-of-frame fusion; Table S11).

SAMM50-PARVB. *SAMM50* encodes the sorting and assembly machinery (SAM) component 50 homolog; *PARVB* encodes β-parvin. SAM-50 is part of the SAM complex, which has a role in integrating β-barrel proteins into the outer mitochondrial membrane. β-Parvin is an actin-binding protein that associates with focal contacts. Parvin is a key regulator of integrin-linked kinase (ILK) and of its downstream pathways. The encoded fusion protein contains an almost entire SAM-50, which only misses its last 15 C-terminal amino acids (Table S11). However, this may lead to disruption of the second major functionally relevant domain of SAM-50. Only two amino acids are contributed by the β-parvin mRNA, as an out-of-frame sequence (Table S11).

URB1-C21orf45. *URB1* encodes the pre-ribosomal–associated protein 1 (Npa1p); *C21orf45* encodes the kinetochore protein homolog A. Npa1p is a component of pre-60S ribosomal particles and associates with small nucleolar ribonucleoprotein particles (RNPs) that are required for peptidyl transferase center modification. The kinetochore protein homolog A is involved in mitosis and associates with chromatin. It also associates with centromeres in interphase cells, from late anaphase to G_1. The fusion protein keeps essentially all the exons of the N-terminal partner (38/39), including its S1385 phosphorylation site (Table S11), suggesting a largely unaltered function. Albeit the C-terminal sequence is not in its native frame, it is unusually long (121 amino acids; Table S11) and may carry novel associated functions.

Chimeras Contain Positive and Negative Regulators of Cell Growth

The six chimeras that were found to be expressed by target cell lines in culture were assayed for a role in cell growth. siRNA targeting the chimeric joint (Table S11) were used to inhibit the expression of the corresponding chimeras in breast cancer cells (Figure 3). Transfection efficiency was optimized by using a co-transfected pEYFP reporter plasmid (Figure 3*C*, *left*); the vast majority of target cells appeared successfully transfected (Figure 3*C*). Transfected siRNA downregulated mRNA chimera levels by ≈75% (Figures 3*B* and S3). To ensure absence of off-target effects, due to artifactual reduction of chimera partner transcript levels, 5′ and 3′ chimeric partners were analyzed in parallel. Levels of partner transcripts of growth regulatory chimeras remained unaffected by siRNA targeting the chimera junction regions (Figure 3*B*).

Remarkably, five of six tested chimeras appeared to regulate cell growth. The strongest growth inhibition in HBL-100 cancer cells

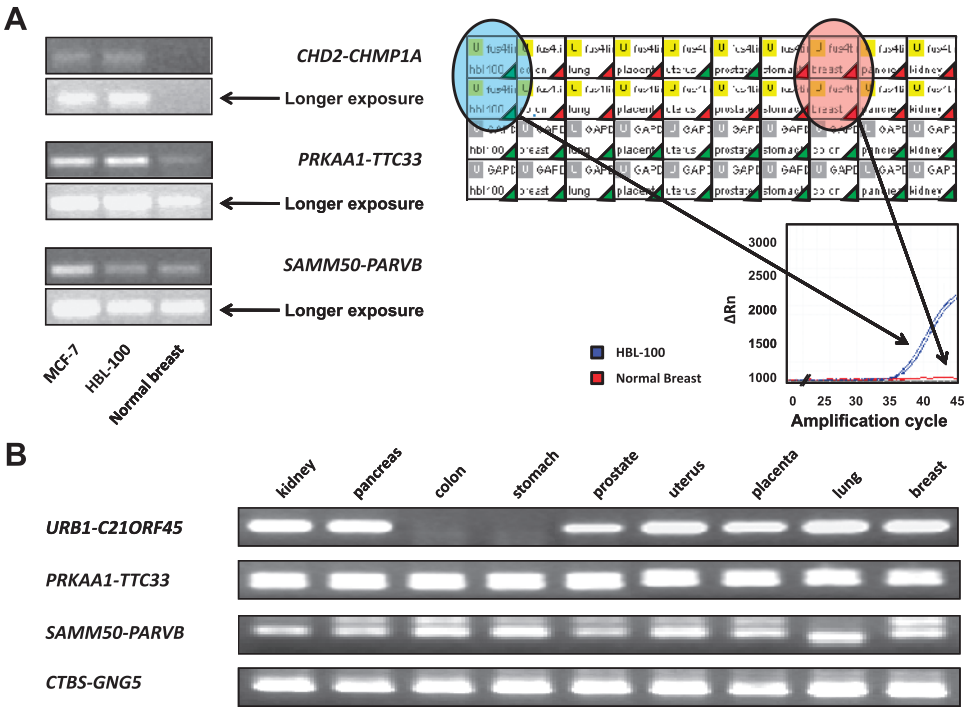

Figure 4. Expression of mRNA chimeras in normal tissues. (A) (Left) Agarose electrophoresis analysis of real-time PCR products (Sybr Green real-time PCR). (Right) Real-time PCR analysis of the *CHD2-CHMP1A* chimeric mRNA in normal tissues (PrimeTime real-time PCR). HBL-100 cell cDNA was used as a positive control. (Bottom) Amplification plots of *CHD2-CHMP1A* mRNA in the HBL-100 breast cancer cell line (blue) and normal breast tissue (red). *CHD2-CHMP1A* (top lanes) and GAPDH (bottom lanes) were amplified in duplicate. Green triangles, successful amplification; red triangles, no amplification. (B) Agarose electrophoresis of nested PCR products of *PRKAA1-TTC33*, *SAMM50-PARVB*, *CTBS-GNG5*, and *URB1-C21orf45* in normal tissues.

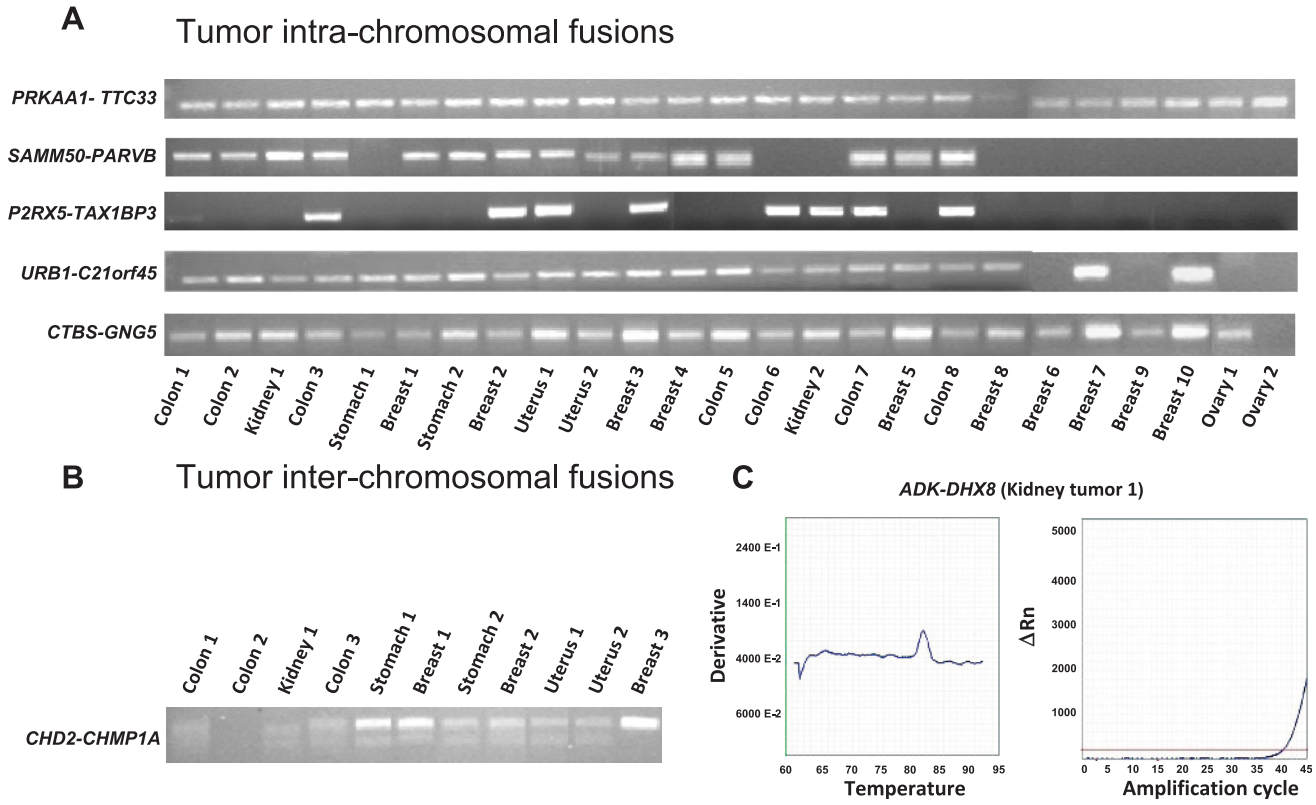

Figure 5. Expression of mRNA chimeras in primary tumors. (A, B) Agarose electrophoresis analysis of amplification products. Tumor origin and sample numbers are indicated. (A) Intrachromosomal chimeras as analyzed by real-time PCR. *PRKAA1-TTC33* and *CTBS-GNG5* were diagnosed in all 25 tumors; *SAMM50-PARVB* chimeric was found in 15 tumors, *P2RX5-TAX1BP3* in 8 tumors, and *URB1-C21orf45* in 21 tumors. (B) PCR amplicons of the interchromosomal *CHD2-CHMP1A* chimera. (C) Interchromosomal *ADK-DHX8* chimeric. Melting temperature and real-time amplification curves.

was caused by down-regulation of *CHD2-CHMP1A* (Figure 3*A*). Parallel growth blockade in MCF-7 cells was observed on shutdown of *PRKAA-TTC33* and *SAMM50-PARVB* (Figures 3*A* and S3B). Monitoring of cell growth inhibition by *PRKAA-TTC33* and *SAMM50-PARVB* siRNA through optical microscopy (Figure 3*D*) and image analysis (Figure S3, B and C) confirmed a dramatic reduction of MCF-7 cell growth. Growth inhibition by *PRKAA-TTC33* and *SAMM50-PARVB* down-regulation was also demonstrated for HBL-100 cells.

We then went on to test *URB1-C21ORF45*–targeting siRNA in ovarian cell lines. Unexpectedly, an increase in cell growth was reproducibly observed in OVCAR-3 (Figure 3*A*) and IGROV-1 cells, which indicates a growth inhibitory role of the *URB1-C21ORF45* chimera. Albeit *URB1-C21ORF45* is expressed by SKOV-3 and HBL-100 cells, the corresponding siRNA had no effects on these cells, suggesting a cell-specific function of these growth inhibitory chimera (Figure 3*A*). These tests were repeated using *CTBS-GNG5*–targeted siRNA. These assays showed that the *CTBS-GNG5* chimera also has a growth inhibitory function in OVCAR-3 and IGROV-1 cells (Figures 3 and S3B). Again, SKOV-3 and HBL-100 cancer cells were insensitive to the inhibitory function of *CTBS-GNG5*, consistent with a differential tuning of chimera-dependent growth-control circuitries in specific cell lines.

Protein-encoding reading frames of the growth regulatory chimeras were analyzed (Table S11). In all cases but one, the downstream partners did not provide in-frame sequences, generating out-of-frame, mostly short chimeric tails. This suggested altered regulation and/or dominant-negative function of a truncated molecule as a mechanism of action of these chimeric products. However, the *CTBS-GNG5* is an in-frame chimera that retains the first 319 amino acids from the

Table 1. Expression of Chimeric mRNA by Tumor Type.

Chimera	Breast* [*n*/10 (%)]	Ovary [*n*/2 (%)]	Stomach [*n*/2 (%)]	Colon [*n*/7 (%)]	Kidney [*n*/2 (%)]	Uterus [*n*/2 (%)]
PRKAA-TTC33	8 (80)	2 (100)	2 (100)	7 (100)	2 (100)	2 (100)
SAMM50-PARVB	5 (50)	–	1 (50)	5 (71)	2 (100)	1 (50)
P2RX5-TAX1BP3	2 (20)	–	–	5 (71)	1 (50)	1 (50)
URB1-C21ORF45	8 (80)	–	1 (100)	7 (100)	2 (100)	2 (100)
CTBS-GNG5	8 (80)	1 (50)	2 (100)	7 (100)	2 (100)	2 (100)
CHD2-CHMP1A	3 (30)	–	2 (100)	2 (28)	1 (50)	2 (100)
ADK-DHX8	–	–	–	1 (14)	1 (50)	–

–, Not detected.
*Tumors; total numbers are below each histotype.

N-terminal chitobiase and most of the C-terminal Gγ5 (41 amino acids), including its Gβ-binding interface (Figure S2). This suggested that the chimeric protein can bind its Gβ partner in trimeric G proteins (Supplemental Sequence Data).

Chimera Expression in Normal Tissues

We assessed the presence and expression levels of the five growth-controlling chimeras in mRNA from normal tissues (breast, lung, placenta, uterus, prostate, stomach, colon, pancreas, and kidney) by nested and real-time PCR. The four intrachromosomal chimeras (*PRKAA-TTC33*, *SAMM50-PARVB*, *URB1-C21ORF45*, and *CTBS-GNG5*) were detected in all screened normal tissues (Figure 4). This was consistent with previous findings on the expression of oncogenic mRNA chimeras in normal tissues [4,13,14,17–20]. However, we found essentially no trace of the *CHD2-CHMP1A* interchromosomal chimera in normal tissues. *CHD2-CHMP1A* was expressed by almost all cancer cell lines (13/14), thus appearing as a cancer-related event.

Expression of Growth Regulatory Chimeras in Primary Tumors

In vitro cell growth regulatory chimeras are expressed by different cancer histotypes. Total RNA was extracted from breast, ovarian, gastric, colon, kidney, and uterine tumors [13,48], was reverse transcribed, and amplified. We took advantage of chimeric-band melting-temperature specificity peaks (Figure S3E) to select for *bona fide* amplification candidates. Amplified candidates were then systematically sequenced. *PRKAA-TTC33* was detected in all 25 of these tumors, *SAMM50-PARVB* in 15 tumors, *P2RX5-TAX1BP3* in 8 tumors, and *URB1-C21ORF45* in 21 tumors; *CTBS-GNG5* was detected in almost all tumors (Figure 5*A* and Table 1); *CHD2-CHMP1A* was identified in 11 tumors (Figure 5*B*). *ADK-DHX8* was diagnosed in two tumors (Figure 5*C*). Hence, growth regulatory chimeras are broadly expressed in human tumors but in heterogeneous manners. This suggests a positive selective pressure [49] for a fusion mRNA-based growth regulatory mechanism during tumor development, which appears to operate in a chimera and tumor-type–specific manner.

Discussion

We have opened the field of the *in silico* identification of mRNA chimeras in cancer cells, through analysis of cDNA sequence databanks [24]. NGS approaches have enormously increased the amount of sequencing data of potential use for chimera discovery. However, short-read second-generation NGS analyses identify mRNA chimeras through a probabilistic fitting of highly multiplexed short-tag data sets [7,19,20,25–28,50–53], which severely affects both specificity and sensitivity of detection of mRNA chimeras. However, rapid progress is being made toward achieving longer sequence reads and higher sequencing accuracy, which allows to reduce sequence errors while improving contig assembly procedures. To permit high-throughput, high-specificity chimera discovery in long-read sequence data sets, we have developed the FusionMiner search strategy. This was shown to reach a 95.9% chimera identification specificity, with a low 4.1% false-negative classification rate. This search strategy was extensively validated by RT-PCR and cDNA sequencing (Table S1b).

Global chimera frequencies were computed for separate sequencing projects. Analysis of a human transcriptomic 454 data set of 19,527 contigs and 173,005 singletons led to the identification of four sequences as *bona fide* chimeras, for a chimera frequency of 4/192,532, i.e., 2×10^{-5} of all mRNA. High-throughput sequencing of cDNA libraries from tumors and corresponding normal tissues generated 1,774,453 long-read sequences. Twenty-five were identified by FusionMiner as *bona fide* chimeras, for a chimera frequency of 25/1,774,453, i.e., 1.4×10^{-5}, in remarkable agreement with the NGS data. Taken together, these findings suggest a chimera frequency of $\approx 2 \times 10^{-5}$ in cellular transcriptomes. Issues of data set size and of transcriptome tissue specificity suggest these to be minimal estimates. A proof of principle of this scenario was obtained, as one of the interchromosomal chimeras, which could not be detected in cell lines, and was identified in 2 of 10 primary breast cancers.

Most of the chimeras analyzed were shown to have a regulatory role in transformed cell growth [54,55]. Notably, tumor growth inhibitory mRNA chimeras, e.g., *URB1-C21ORF45* and *CTBS-GNG5*, were also discovered. Of interest, these were shown to have inhibitory capacity on the growth of a subset of ovarian cancer cells, whereas other ovarian and breast cancer cells were not affected, suggesting different regulatory contexts for chimera-driven growth control in different cell lines. Most tumors were shown to express these growth regulatory chimeras, consistent with a positive selective pressure for exploiting this growth regulatory mechanism during tumor development.

Acknowledgments

We thank M. Iacono for providing the Roche NGS data sets and C. Berrie for critical reading and editing of the manuscript.

References

[1] Mitelman F, Johansson B, and Mertens F (2007). The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* **7**, 233–245.

[2] Mercer TR, Dinger ME, and Mattick JS (2009). Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10**, 155–159.

[3] Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648.

[4] Rickman DS, Pflueger D, Moss B, VanDoren VE, Chen CX, de la Taille A, Kuefer R, Tewari AK, Setlur SR, Demichelis F, et al. (2009). SLC45A3-ELK4 is a novel and frequent erythroblast transformation–specific fusion transcript in prostate cancer. *Cancer Res* **69**, 2734–2738.

[5] Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, et al. (2007). Identification of the transforming *EML4-ALK* fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566.

[6] Fehr A, Roser K, Heidorn K, Hallas C, Loning T, and Bullerdiek J (2008). A new type of MAML2 fusion in mucoepidermoid carcinoma. *Genes Chromosomes Cancer* **47**, 203–206.

[7] Palanisamy N, Ateeq B, Kalyana-Sundaram S, Pflueger D, Ramnarayanan K, Shankar S, Han B, Cao Q, Cao X, Suleman K, et al. (2010). Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat Med* **16**, 793–798.

[8] Ciampi R, Knauf JA, Kerler R, Gandhi M, Zhu Z, Nikiforova MN, Rabes HM, Fagin JA, and Nikiforov YE (2005). Oncogenic AKAP9-BRAF fusion is a novel mechanism of MAPK pathway activation in thyroid cancer. *J Clin Invest* **115**, 94–101.

[9] Santoro M, Melillo RM, and Fusco A (2006). RET/PTC activation in papillary thyroid carcinoma: European Journal of Endocrinology Prize Lecture. *Eur J Endocrinol* **155**, 645–653.

[10] Edwards PA (2010). Fusion genes and chromosome translocations in the common epithelial cancers. *J Pathol* **220**, 244–254.

[11] Skotheim RI, Thomassen GO, Eken M, Lind GE, Micci F, Ribeiro FR, Cerveira N, Teixeira MR, Heim S, Rognes T, et al. (2009). A universal assay for detection of oncogenic fusion transcripts by oligo microarray analysis. *Mol Cancer* **8**, 5.

[12] Bruzik JP and Maniatis T (1992). Spliced leader RNAs from lower eukaryotes are trans-spliced in mammalian cells. *Nature* **360**, 692–695.

[13] Guerra E, Trerotola M, Dell'Arciprete R, Bonasera V, Palombo B, El-Sewedy T, Ciccimarra T, Crescenzi C, Lorenzini F, Rossi C, et al. (2008). A bicistronic

CYCLIN D1-TROP2 mRNA chimera demonstrates a novel oncogenic mechanism in human cancer. *Cancer Res* **68**, 8113–8121.
[14] Li H, Wang J, Mor G, and Sklar J (2008). A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science* **321**, 1357–1361.
[15] Terrinoni A, Dell'Arciprete R, Fornaro M, Stella M, and Alberti S (2001). Cyclin D1 gene contains a cryptic promoter that is functional in human cancer cells. *Genes Chromosomes Cancer* **31**, 209–220.
[16] Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, and Chinnaiyan AM (2009). Transcriptome sequencing to detect gene fusions in cancer. *Nature* **458**, 97–101.
[17] Communi D, Suarez-Huerta N, Dussossoy D, Savi P, and Boeynaems J-M (2001). Cotranscription and intergenic splicing of human $P2Y_{11}$ and SSF_1 genes. *J Biol Chem* **276**, 16561–16566.
[18] Li H, Wang J, Ma X, and Sklar J (2009). Gene fusions and RNA trans-splicing in normal and neoplastic human cells. *Cell Cycle* **8**, 218–222.
[19] Pflueger D, Terry S, Sboner A, Habegger L, Esgueva R, Lin PC, Svensson MA, Kitabayashi N, Moss BJ, Macdonald TY, et al. (2010). Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. *Genome Res* **21**, 56–67.
[20] Edgren H, Murumagi A, Kangaspeska S, Nicorici D, Hongisto V, Kleivi K, Rye IH, Nyberg S, Wolf M, Borresen-Dale AL, et al. (2011). Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol* **12**, R6.
[21] Ambrogi F, Biganzoli E, Querzoli P, Ferretti S, Boracchi P, Alberti S, Marubini E, and Nenci I (2006). Molecular subtyping of breast cancer from traditional tumor marker profiles using parallel clustering methods. *Clin Cancer Res* **12**, 781–790.
[22] Rabbitts TH and Stocks MR (2003). Chromosomal translocation products engender new intracellular therapeutic technologies. *Nat Med* **9**, 383–386.
[23] Cimoli G, Malacarne D, Ponassi R, Valenti M, Alberti S, and Parodi S (2004). Meta-analysis of the role of p53 status in isogenic systems tested for sensitivity to cytotoxic antineoplastic drugs. *Biochim Biophys Acta* **1705**, 103–120.
[24] Romani A, Guerra M, Trerotola M, and Alberti S (2003). Detection and analysis of spliced chimeric mRNAs in sequence databanks. *Nucleic Acids Res* **31**, 1–8.
[25] Sboner A, Habegger L, Pflueger D, Terry S, Chen DZ, Rozowsky JS, Tewari AK, Kitabayashi N, Moss BJ, Chee MS, et al. (2010). FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. *Genome Biol* **11**, R104.
[26] Nacu S, Yuan W, Kan Z, Bhatt D, Rivers CS, Stinson J, Peters BA, Modrusan Z, Jung K, Seshagiri S, et al. (2011). Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. *BMC Med Genomics* **4**, 11.
[27] Asmann YW, Hossain A, Necela BM, Middha S, Kalari KR, Sun Z, Chai HS, Williamson DW, Radisky D, Schroth GP, et al. (2011). A novel bioinformatics pipeline for identification and characterization of fusion transcripts in breast cancer and normal cell lines. *Nucleic Acids Res* **39**, e100.
[28] Iyer MK, Chinnaiyan AM, and Maher CA (2011). ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics* **27**, 2903–2904.
[29] Carletti E, Guerra E, and Alberti S (2006). The forgotten variables of DNA array hybridization. *Trends Biotechnol* **24**, 443–448.
[30] Shendure J and Ji H (2008). Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135–1145.
[31] Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, and Pallen MJ (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* **30**, 434–439.
[32] Mardis ER (2011). A decade's perspective on DNA sequencing technology. *Nature* **470**, 198–203.
[33] Orsulic S, Li Y, Soslow RA, Vitale-Crosss LA, Gutkind JS, and Varmus HE (2002). Induction of ovarian cancer by defined multiple genetic changes in a mouse model system. *Cancer Cell* **1**, 53–62.
[34] Alberti S and Fornaro M (1990). Higher transfection efficiency of genomic DNA purified with a guanidinium thiocyanate-based procedure. *Nucleic Acids Res* **18**, 351–353.
[35] Dell'Arciprete R, Stella M, Fornaro M, Ciccocioppo R, Capri MG, Naglieri AM, and Alberti S (1996). High-efficiency expression gene cloning by flow cytometry. *J Histochem Cytochem* **44**, 629–640.
[36] Alberti S, Nutini M, and Herzenberg LA (1994). DNA methylation prevents the amplification of TROP1, a tumor-associated cell surface antigen gene. *Proc Natl Acad Sci USA* **91**, 5833–5837.
[37] Alberti S and Herzenberg LA (1988). DNA methylation prevents transfection of genes for specific surface antigens. *Proc Natl Acad Sci USA* **85**, 8391–8394.
[38] Alberti S, Parks DR, and Herzenberg LA (1987). A single laser method for subtraction of cell autofluorescence in flow cytometry. *Cytometry* **8**, 114–119.
[39] Biganzoli E, Coradini D, Ambrogi F, Garibaldi JM, Lisboa P, Soria D, Green AR, Pedriali M, Piantelli M, Querzoli P, et al. (2011). p53 status identifies two subgroups of triple-negative breast cancers with distinct biological features. *Jpn J Clin Oncol* **41**, 172–179.
[40] Querzoli P, Coradini D, Pedriali M, Boracchi P, Ambrogi F, Raimondi E, La Sorda R, Lattanzio R, Rinaldi R, Lunardi M, et al. (2010). An immunohistochemically positive E-cadherin status is not always predictive for a good prognosis in human breast cancer. *Br J Cancer* **103**, 1835–1839.
[41] Tinari N, Lattanzio R, Natoli C, Cianchetti E, Angelucci D, Ricevuto E, Ficorella C, Marchetti P, Alberti S, Piantelli M, et al. (2006). Changes of topoisomerase IIα expression in breast tumors after neoadjuvant chemotherapy predicts relapse-free survival. *Clin Cancer Res* **12**, 1501–1506.
[42] Brummelkamp TR, Bernards R, and Agami R (2002). A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553.
[43] Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, and Tuschl T (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498.
[44] Semizarov D, Frost L, Sarthy A, Kroeger P, Halbert DN, and Fesik SW (2003). Specificity of short interfering RNA determined through gene expression signatures. *Proc Natl Acad Sci USA* **100**, 6347–6352.
[45] Chalk AM, Wahlestedt C, and Sonnhammer EL (2004). Improved and automated prediction of effective siRNA. *Biochem Biophys Res Commun* **319**, 264–274.
[46] Bonasera V, Alberti S, and Sacchetti A (2007). Protocol for high-sensitivity/long linear-range spectrofluorimetric DNA quantification using ethidium bromide. *Biotechniques* **43**, 173–176.
[47] Akiva P, Toporik A, Edelheit S, Peretz Y, Diber A, Shemesh R, Novik A, and Sorek R (2006). Transcription-mediated gene fusion in the human genome. *Genome Res* **16**, 30–36.
[48] Querzoli P, Pedriali M, Rinaldi R, Lombardi AR, Biganzoli E, Boracchi P, Ferretti S, Frasson C, Zanella C, Ghisellini S, et al. (2006). Axillary lymph node nanometastases are prognostic factors for disease-free survival and metastatic relapse in breast cancer patients. *Clin Cancer Res* **12**, 6696–6701.
[49] Alberti S (1997). The origin of the genetic code and protein synthesis. *J Mol Evol* **45**, 352–358.
[50] Ge H, Liu K, Juan T, Fang F, Newman M, and Hoeck W (2011). FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics* **27**, 1922–1928.
[51] Hu Y, Wang K, He X, Chiang DY, Prins JF, and Liu J (2010). A probabilistic framework for aligning paired-end RNA-seq data. *Bioinformatics* **26**, 1950–1957.
[52] McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MG, Griffith M, Heravi Moussavi A, Senz J, Melnyk N, et al. (2011). deFuse: an algorithm for gene fusion discovery in tumor RNA-seq data. *PLoS Comput Biol* **7**, e1001138.
[53] Kinsella M, Harismendy O, Nakano M, Frazer KA, and Bafna V (2011). Sensitive gene fusion detection using ambiguously mapping RNA-Seq read pairs. *Bioinformatics* **27**, 1068–1075.
[54] Guerra E, Trerotola M, Aloisi AL, Tripaldi R, Vacca G, La Sorda R, Lattanzio R, Piantelli M, and Alberti S (2012). The Trop-2 signalling network in cancer growth. *Oncogene.* DOI: 10.1038/onc.2012.151 [E-pub ahead of print].
[55] Trerotola M, Cantanelli P, Guerra E, Tripaldi R, Aloisi AL, Bonasera V, Lattanzio R, de Lange R, Weidle UH, Piantelli M, et al. (2012). Upregulation of Trop-2 quantitatively stimulates human cancer growth. *Oncogene.* DOI: 10.1038/onc.2012.36 [E-pub ahead of print].

Da: Editor <edassist@neoplasia.com>
Data: 30 settembre 2012 21:49:46 GMT+02:00
A: s.alberti@unich.it
Oggetto: **Decision Made MS # 12-1342 Version 1**

Dear Dr. Saverio Alberti,

Your manuscript entitled "Long-range transcriptome sequencing reveals cancer cell growth regulatory chimeric mRNAs", MS No. 12-1342, was presented to the editorial board at this weeks meeting.

After summarizing the key findings with the findings with the editorial board, I am pleased to inform you that your manuscript has been accepted for publication without revision. The manuscript will be sent to the publisher immediately for inclusion in the next available issue. Despite the strength of the key findings, i would have under normal circumstances preferred a critique, but at the risk of delaying publication of the work. Therefore I would like to accept the paper and ask that you feel free to modify the manuscript as you feel fit in the next couple of days before it is sent to the printers.

Thank you for considering pub lishing this important work in Neoplasia. Congratulations, and please keep us in mind for your next submission.

Sincerely,

Dr. Al Rehemtulla
Editor, Neoplasia
Neoplasia Press