Integrated Systems and Technologies

Cancer Research

Exon Array Analyses across the NCI-60 Reveal Potential Regulation of TOP1 by Transcription Pausing at Guanosine Quartets in the First Intron

William C. Reinhold[1], Jean-Louis Mergny[3], Hongfang Liu[1,4], Michael Ryan[5], Thomas D. Pfister[6], Robert Kinders[6], Ralph Parchment[6], James Doroshow[1,2], John N. Weinstein[7], and Yves Pommier[1]

Abstract

Because topoisomerase 1 (TOP1) is critical for the relaxation of DNA supercoils and because it is the target for the anticancer activity of camptothecins, we assessed *TOP1* transcript levels in the 60 cell line panel (the NCI-60) of the National Cancer Institute's anticancer drug screen. *TOP1* expression levels varied over a 5.7-fold range across the NCI-60. HCT116 colon and MCF-7 breast cancer cells were the highest expressers; SK-MEL-28 melanoma and HS578T breast carcinoma cells were the lowest. TOP1 mRNA expression was highly correlated with Top1 protein levels, indicating that *TOP1* transcripts could be conveniently used to monitor Top1 protein levels and activity in tissues. Assessment of the *TOP1* locus by array comparative genomic hybridization across the NCI-60 showed copy numbers ranging from 1.71 to 4.13 and a statistically significant correlation with *TOP1* transcript levels ($P < 0.01$). Further analyses of *TOP1* expression on an exon-specific basis revealed that exon 1 expression was generally higher and less variable than expression of the other exons, suggesting some form of transcriptional pausing regulation between exons 1 and 2. Accordingly, we found the presence of multiple evolutionarily conserved potential G-quadruplex–forming sequences in the first *TOP1* intron. Physicochemical tests for actual quadruplex formation by several of those sequences yielded quadruplex formation for two of them and duplex formation for one. The observations reported here suggest the hypothesis that there is a conserved negative transcription regulator within intron 1 of the *TOP1* gene associated with a quadruplex-prone region. *Cancer Res; 70(6); 2191–203. ©2010 AACR.*

Introduction

Topoisomerase 1 (Top1) catalyzes DNA strand breakage through the reversible formation of covalent bonds between an enzyme tyrosyl oxygen and a DNA phosphorus (1). The strand breakage/religation allows relaxation of DNA supercoiling to facilitate the diverse processes of replication, transcription, repair, recombination, and chromatin remodeling (1–3). Top1 is therapeutically important because Top1 inhibitors derived from the plant alkaloid camptothecin (topotecan and irinotecan) are routinely used to treat colon, ovarian, and lung cancers in adults and neuroblastoma and sarcoma in children. Noncamptothecin Top1 inhibitors are in preclinical development (1, 4).

Top1 is expressed ubiquitously in eukaryotic cells, including nonreplicative and postmitotic cells. Expression of the TOP1 gene is essential in animals, and its homozygous disruption is early embryonic lethal (5). Even its underexpression leads to alterations in DNA replication and genomic organization (6). Reduced levels of Top1 may be rate-limiting for the relaxation of positive DNA supercoiling ahead of replication and transcription complexes (1, 6, 7). Accumulation of supercoiling may promote replication fork collapse and transcriptional R-loops (8). TOP1 overexpression is also toxic (9). Too much Top1 may promote the formation of Top1 cleavage complexes at endogenous DNA lesions (abasic sites, mismatches, oxidized bases, nicks, and DNA adducts; refs. 1, 10). Top1 protein levels are also critical for responses to anticancer therapy; cell killing by Top1 inhibitors, including camptothecins and indenoisoquinolines, is positively correlated with Top1 expression (6, 11–15).

The 60 cancer cell lines (the NCI-60) of the National Cancer Institute (NCI) Developmental Therapeutics anticancer screen exhibit differential expression patterns, which is an

Authors' Affiliations: [1]Laboratory of Molecular Pharmacology and [2]Developmental Therapeutics Program, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland; [3]Institut National de la Sante et de la Recherche Medicale U565, Muséum National d'Histoire Naturelle USM 503, Regulation et Dynamique des Genomes (Laboratoire de Biophysique), Paris, France; [4]Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, District of Columbia; [5]Tiger Team Consulting, Fairfax, Virginia; [6]Laboratory of Human Toxicology and Pharmacology, Applied/Developmental Research Support Directorate, Science Applications International Corporation-Frederick, Inc., National Cancer Institute-Frederick, Frederick, Maryland; and [7]Department of Bioinformatics and Computational Biology and Department of Systems Biology, M.D. Anderson Cancer Center, Houston, Texas

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Author: William C. Reinhold, NIH, 9000 Rockville Pike, Building 37, Room 5056, Bethesda, MD 20894. Phone: 301-496-9571; Fax: 301-402-0752; E-mail: wcr@mail.nih.gov.

doi: 10.1158/0008-5472.CAN-09-3528

asset in identifying relationships to general processes, such as alterations in DNA copy number and pharmacologic response (16, 17). In a recent study, we showed that Top1 expression varies significantly across the NCI-60 and that Top1 protein expression is significantly correlated with *TOP1* mRNA expression across those cells (18). The NCI-60 panel (19, 20) constitutes a unique database for pharmacologic inquiry, because it has been characterized more extensively at the DNA, RNA, protein, chromosomal, functional, and pharmacologic levels than any other set of cells in existence[8] (17, 20–24).

The aim of the present study was to take advantage of our multifaceted molecular profiling of the NCI-60 to elucidate genetic mechanism(s) that regulate *TOP1* transcript expression. For that purpose, we integrated data on (*a*) *TOP1* transcript expression, using data from three different platforms, Human Genome U95 (HG-U95; ref. 24), Human Genome U-133 (HG-U133; ref. 24), and Human Genome U133 Plus 2.0 (HG-U133 Plus 2.0); (*b*) *TOP1* transcript expression at the exon level for the 21 *TOP1* exons using Affymetrix GH Exon 1.0 ST Arrays; (*c*) TOP1 protein expression (18); and (*d*) DNA copy number, by array comparative genomic hybridization (aCGH) using NimbleGen arrays from the NCI-60 screen. Those joint profiling data generated a hypothesis about the regulation of TOP1 expression that we then followed up with biochemical and biophysical studies.

Materials and Methods

Quantitation of Top1 transcript expression using three microarray platforms. As described previously (18), transcript expression data for the NCI-60 were obtained from our Affymetrix Human Genome U95 Set (HG-U95; ~60,000 features; Affymetrix, Inc.) and Human Genome U133 (HG-U133a and b; ~44,000 feature; ref. 24). We also analyzed data for *TOP1* from our Human Genome U133 Plus 2.0 Arrays (HG-U133 Plus 2.0; ~47,000 features). We used GC robust multi-array average (RMA) normalization for the HG-U95 and HG-U133. RMA was used for the HG-U133 Plus 2.0 (25). To be included in the calculation of *TOP1* expression levels (Fig. 1A), probe sets were required to have an intensity range of ≥1.2 $\log_2$ and be consistent to the pattern of expression of the other probe sets, using a Pearson's correlation coefficient cutoff of ≥0.52. Probes that passed these quality controls (Fig. 1A) were used to calculate the *z* score[9] average for *TOP1* by first determining the *z* score for each probe by subtracting the 60-cell mean and dividing by the SD (the second to last and last rows of numbers in Fig. 1A, respectively) and then averaging the resulting values for each cell line. Data from the HG-U95 and HG-U133 microarrays can be accessed at our relational database, CellMiner.[8]

The sequence within *TOP1* that the individual probes (that make up the probe sets) from the HG-U95, HG-U133, and HG-U133 Plus 2.0 arrays hybridize to were visualized (Fig. 1C) using SpliceCenter.[10] SpliceCenter is a suite of user-friendly tools that evaluate the effect of gene splice variation on a variety of molecular biological techniques.

Exon-specific Top1 transcript expression. Transcript expression data for each of the 21 *TOP1* exons were obtained using the Affymetrix GeneChip Human Exon 1.0 ST (GH Exon 1.0 ST) Array according to manufacturer's instructions (chips for the studies from John Ward, Thomas Gingeras, and colleagues at Affymetrix). Assays were run at GeneLogic (under the aegis of E. Kaldjian) following manufacturer's recommendations. Technical triplicate arrays were hybridized for each cell line, except for renal UO-31, which was done in duplicate, and central nervous system (CNS) SF-539, which was missing. Transcript data were normalized by RMA (25) using Partek Genomics Suite version 6.3.

To calculate mean-centered intensities across cell lines for each probe set, the technical triplicates for each cell line were averaged for each probe set. All probe sets were then quality-controlled for either being at background levels for all cell lines (and thus uninformative about pattern) or being "dead" (unresponsive, although not at background). Probes that failed either of those criteria were dropped from further analysis. When there were multiple probe sets for a single exon, averages were taken next to yield a single value for each exon.

Data for each exon were next mean-centered across the 59 cell line intensities available (CNS SF-539 was unavailable). The mean-centered intensity values (*M*) for each exon were calculated as *M* = intensity value for the exon − (average of intensity values for the exon across 59 cell lines).

Visual representations of the TOP1 introns, exons, and probe locations were generated using SpliceCenter[10] and the clustered image map (Fig. 3A) generated by CIMminer.[11]

The statistical significance ($P = 2.2 \times 10^{-16}$) of differences in intensity between *TOP1* exon 1 and exons 2 to 21 was calculated by Welch's *t* test on the basis of measurements using 182 microarrays. On each array, there were four probes specific for exon 1 (yielding 4 × 182 = 712 measurements) and 33 probes specific for exons 2 to 21 (yielding 33 × 182 = 6,006 measurements).

Exon and intron 1 nucleotide information and G-quadruplex identification. Human, mouse, and rat sequence information was obtained from National Center for Biotechnology Information.[12] For Fig. 3A, the exon sizes were derived from information obtained from locus NM_003286. The TOP1 intron 1 sizes and sequences were obtained from locus NC_000020, NC_000068, and NC_005102 for human, mouse, and rat, respectively. Fly sequences were obtained from FlyBase.[13] The complementary strand sequence was generated using the Molecular Biologist's Workbench's Data

[8] See http://discover.nci.nih.gov.
[9] See http://en.wikipedia.org/wiki/Standard_score.
[10] See http://discover.nci.nih.gov/splicecenter.
[11] See http://discover.nci.nih.gov/cimminer/.
[12] See http://www.ncbi.nlm.nih.gov/.
[13] See http://flybase.org/reports/FBgn0004924.html.
[14] See http://www.bioinformatics.org/JaMBW/.

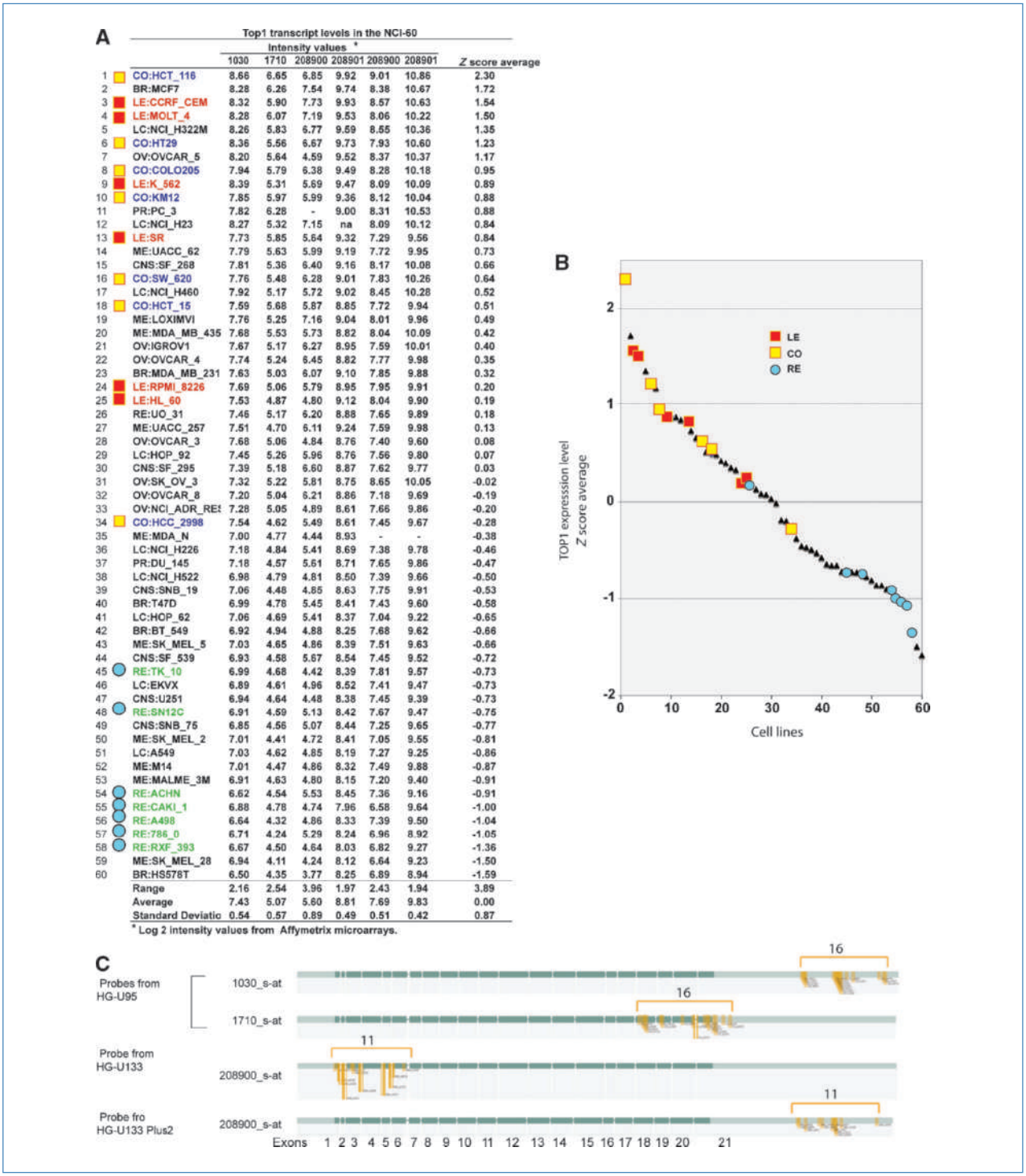

Figure 1. *TOP1* transcript level intensities, *z* score distribution in the NCI-60, and exon location of probe sets. A, table of intensity levels for six probe sets from the HG-U95, HG-U133, and HG-U133 Plus 2.0 microarrays with their combined *z* score means (presented in descending order). The colon (CO), leukemia (LE), and renal (RE) tissue-of-origin lines are represented in yellow, red, and blue, respectively. The numerical series (1–60) for the cell lines corresponds to the 1-to-60 *x* axis in B. B, scatter plot depicting the distribution of *TOP1* transcript levels across the NCI-60. The 60 cell lines are ordered on the *x* axis by decreasing *z* score mean *TOP1* expression (see A). C, location of the probes within each probe set used to measure *TOP1* intensity (see A). The exons of *TOP1* from reference sequence NM_003286 are visualized using SpliceCenter's Array-Check program. The probe set designations (e.g., 1030_s_at) appear on the left. The gold stripes indicate the individual probes that make up each probe set. The number of probes in each probe set is indicated in the above brackets (16 for 1030_s_at and 1710_s_at and 11 for 208900_s_at and 208901_s_at). The gray and green boxes indicate untranslated and translated regions of the *TOP1* gene, respectively.

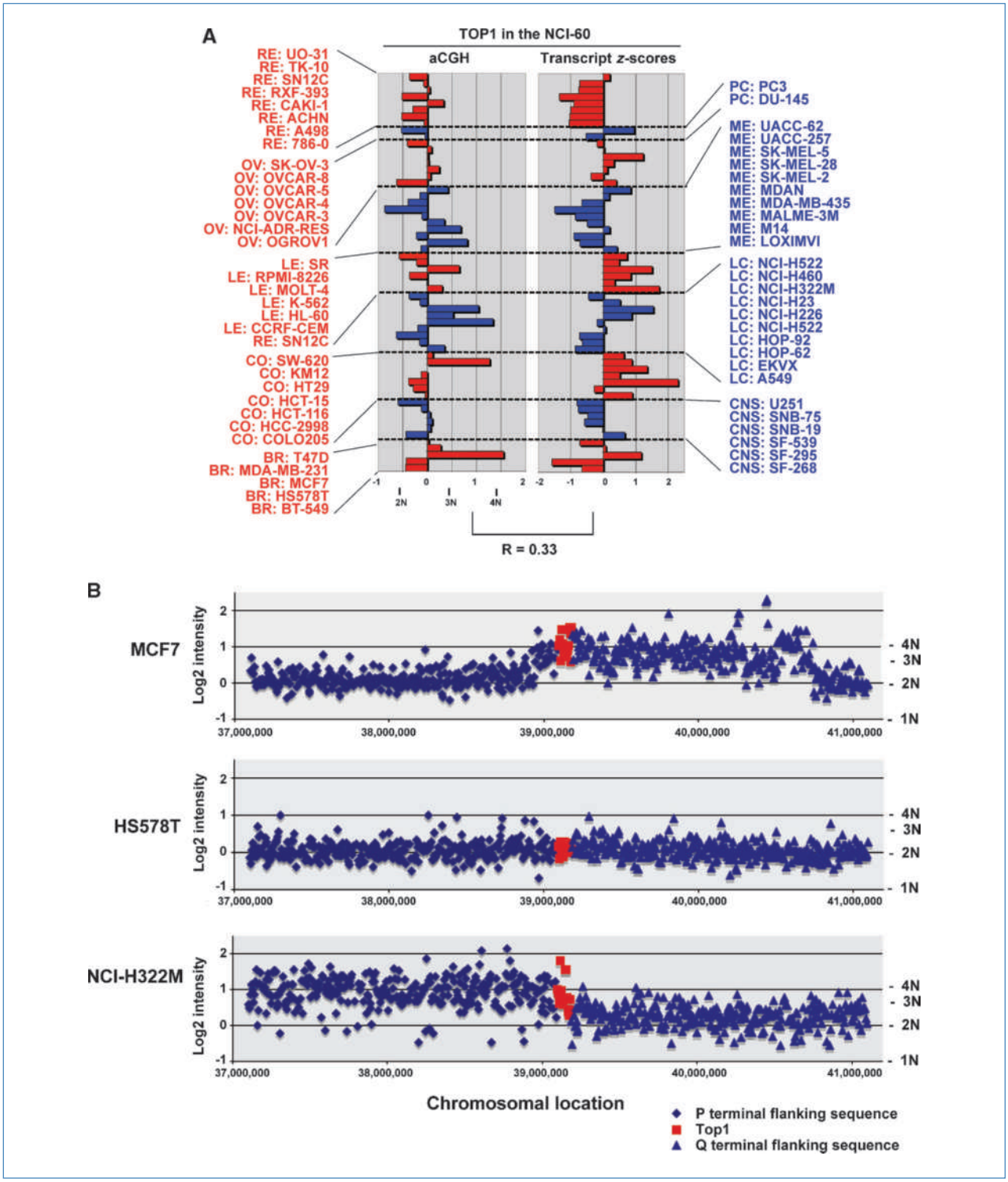

Figure 2. Estimated *TOP1* DNA copy numbers and transcript (Fig. 1A) levels across the NCI-60 and correlation between the two. A, mean-centered aCGH intensities and estimated DNA copy numbers are plotted on the left-hand panel *x* axis, and transcript expression *z* scores (see Fig. 1A) are plotted on the right-hand panel *x* axis. For the aCGH, 2N, 3N, and 4N are the DNA copy numbers. For both panels, the *y* axis corresponds to cell lines, grouped by tissue of origin. *R* is the Pearson's correlation comparison of the two. B, detailed view of *TOP1* aCGH $\log_2$ intensity values from NimbleGen 385k aCGH microarray and two megabases of flanking sequence on both sides. The *x* axis is the nucleotide location within chromosome 20. The *y* axis contains two measurements. The $\log_2$ intensity values of the ratio of the cell line DNA compared with normal DNA is on the left side, and the DNA copy number (1N, 2N, 3N, and 4N) is on the right side.

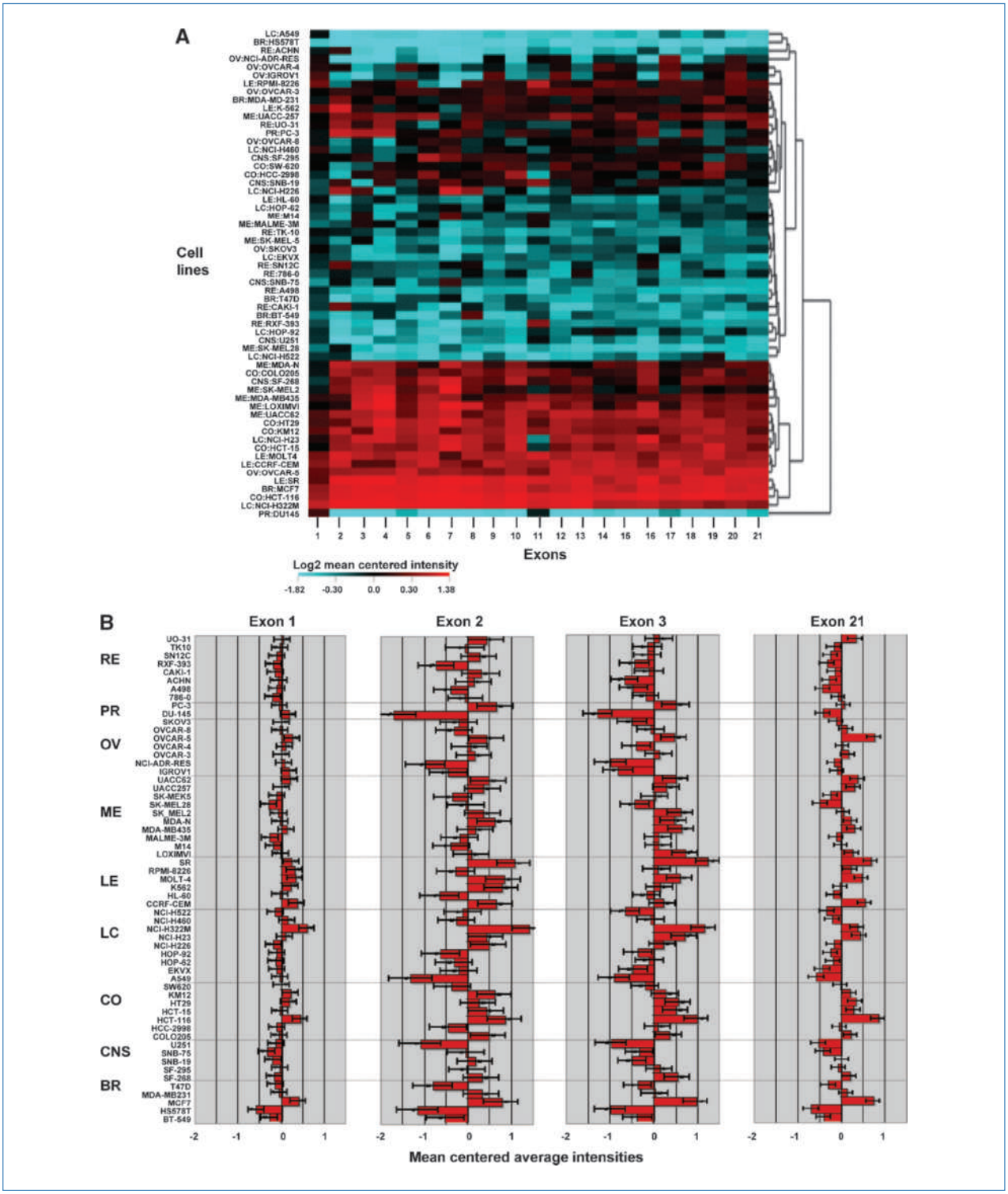

Figure 3. Variation of mean-centered $\log_2$ *TOP1* probe set intensities from Affymetrix GH Exon 1.0 ST microarrays. A, clustered image map (heat map) for mRNA expression for each of the 21 exons of *TOP1* across the NCI-60 cell lines. The cell lines are clustered on the *y* axis based on Euclidean distance using average linkage. The *x* axis displays the 21 *TOP1* exons. The colors are distributed to indicate mean-centered (across the 60 cell lines) $\log_2$ intensity values. Red indicates high, blue indicates low, and black indicates midrange intensity. B, differential *TOP1* transcript levels in exon 1 versus the other *TOP1* exons across the NCI-60. Mean-centered expression intensity of TOP1 transcripts corresponding to exons 1, 2, 3, and 21. In all panels, the *x* axis is mean-centered $\log_2$ intensity level. The cell lines on the *y* axis are grouped by tissue of origin. Error bars indicate 95% confidence intervals derived for technical cell line replicates and, in some instances, multiple probe sets for individual exons.

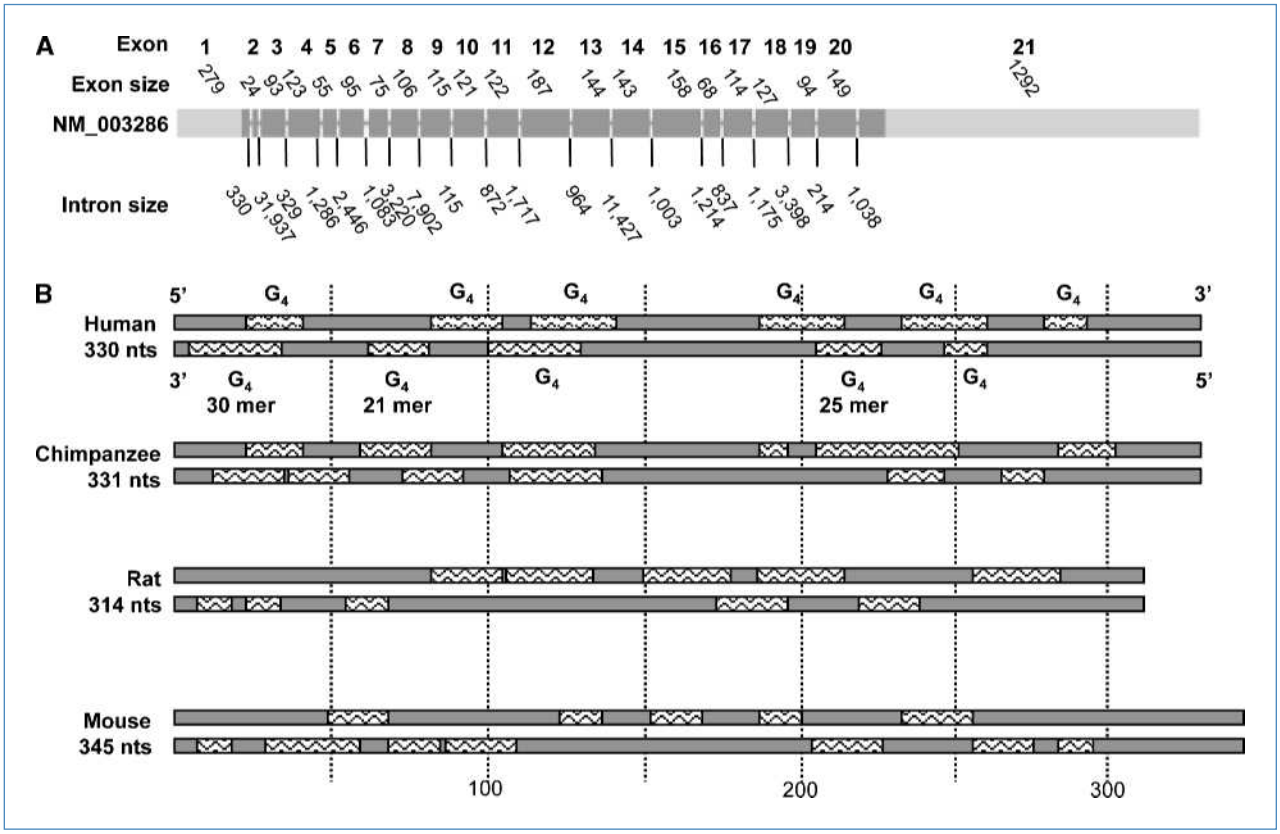

Figure 4. Schematic representation of potential quadruplex-forming G-rich sequences within *TOP1* intron 1. A, *TOP1* exon and intron sizes. The exons are numbered (first row of numbers) and drawn to scale (using SpliceCenter) as reported for NM_003286. The lighter and darker gray regions of NM_003286 indicate untranslated and translated regions, respectively. Sizes of exons and introns are indicated above and below the exon depiction, respectively. B, relative location of the potential quadruplex-forming G-rich sequences for both DNA strands across species. Potential quadruplexes are indicated as lighter blocks (labeled G_4 in human). The locations of the three oligonucleotides tested biochemically for G-quadruplex formation (Fig. 5) are labeled 30-mer, 21-mer, and 25-mer. The vertical dotted lines are spaced at 50-nucleotide intervals to indicate scale for the four species depicted (human, chimpanzee, rat, and mouse).

Manipulation tool.[14] Potential G-quadruplex regions were identified using default conditions at the GRS (quadruplex-forming G-rich sequences) Mapper.[15]

Absorbance and circular dichroism. Oligonucleotides were synthesized by Eurogentec at the 200-nmol scale. Concentrations were estimated using extinction coefficients provided by the manufacturer and calculated using a nearest-neighbor model (26) as described previously. Sequences (Fig. 5A) are given in the 5′ to 3′ direction. Melting experiments were conducted as previously described (27, 28) by recording the absorbance at 295 nm (27, 29). Sequences were tested at least twice at 4-μmol/L strand concentration. Thermal difference spectra (TDS) were obtained from differences between the absorbance spectra from unfolded and folded oligonucleotides (as recorded above and below their T_ms (30). Circular dichroism spectra (CDS) were recorded on a JASCO-810 spectropolarimeter as described previously (28).

aCGH determination of approximate mean TOP1 DNA copy number. The NimbleGen 385,000-feature Human Whole-Genome array (HG17, Build 35) probe microarray[16] yielded data from 15 probes specific for *TOP1*. They are 39095858, 39101405, 39106639, 39111664, 39117552, 39123350, 39128951, 39134754, 39146000, 39154112, 39159375, 39165420, 39170495, 39176079, and 39181834. Approximate mean DNA copy number was calculated as

$$P = C \times L^{\text{intensity}}$$

wherein C = (the correction for generating the intensities as a ratio of the cell line intensity to a normal, 2N, DNA) = 2 and L = log of the intensity values = 2.

HUGO naming of genes. Genes are designated here by their HUGO (Human Genome Organization) names, as promulgated by the Gene Nomenclature Committee.[17]

[15] See http://bioinformatics.ramapo.edu/QGRS/analyze.php.

[16] Varma, in preparation.

[17] See http://www.genenames.org.

Results

TOP1 transcript expression across the NCI-60. To determine relative transcript levels of *TOP1* in the NCI-60, we used six probes from three different microarray platforms: HG-U95 (24), HG-U133 (24), and HG-U133 Plus 2.0. Figure 1A displays these relative levels both as intensity values and as the *z* score average. The use of *z* scores allows comparison of relative levels for data distributions with different means and/or SDs. Values obtained for probes from the three platforms were consistent with each other (mean Pearson's correlation coefficient = 0.72, with a range of 0.52–0.93). The six probes, two of which appear in both HG-U133 and HG-U133 Plus 2.0, hybridize to exons 1 to 6 and 18 to 21 (Fig. 1C), with the majority of probes targeting the 3′ end of the gene. *TOP1* expression in the NCI-60 varied over a 5.7-fold range, within 1 SD of the average transcript variation (a 9.0-fold range) for 26 housekeeping genes (as defined in ref. 31).

Colon HCT116 and breast MCF-7 cells showed the highest TOP1 mRNA levels, and breast HS578T and melanoma SK-MEL-28 showed the lowest (Figs. 1A and 2A). The six leukemia cell lines and six of the seven colon carcinoma cell lines consistently expressed high *TOP1* mRNA (Fig. 1A, B). Their average *z* scores (0.86 and 0.89, respectively) were >3.7 times that of the next highest tissue of origin type, ovarian, at 0.23. HCC-2998 cells were the only colon carcinoma cells with lower than average *TOP1* transcript levels. The breast and prostate lines were the most variable in *TOP1* expression, with SDs of 1.25, and 0.95, respectively. The renal lines were the lowest expressers (Fig. 1A, B), with an average *z* score of −0.83. Breast, CNS, and melanoma formed a second tier of negative expressers, with *z*-score averages of −0.16, −0.35, and −0.34, respectively. TOP1 transcript and protein levels correlated with one another at statistically significant levels ($r = 0.80$, $P < 0.001$; ref. 18).

TOP1 DNA copy number and transcript levels correlate with each other at statistically significant levels across the NCI-60. To evaluate whether DNA copy number contributes to TOP1 expression, we used aCGH to determine average copy number levels for the *TOP1* locus and two megabases of flanking region on both sides of it. Those levels were obtained from our studies using NimbleGen 385,000-feature Human Whole-Genome CGH arrays. Based on the average intensities of 15 tiled probes specific for the TOP1 locus,

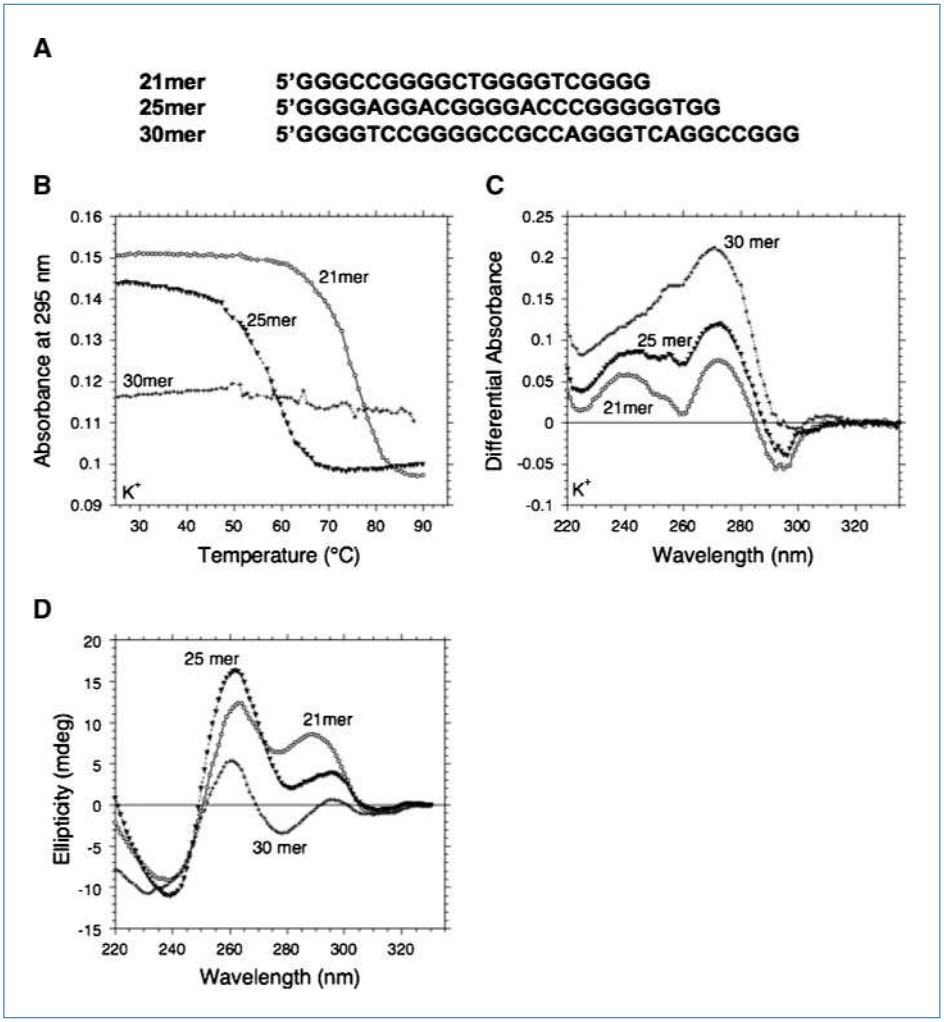

Figure 5. Three oligonucleotides tested for G-quadruplex formation. A, sequences of the three human *TOP1* exon 1 oligonucleotides tested (located in Fig. 4B). B, UV melting profiles at 295 nm in 0.1 mol/L KCl. C, TDS in 0.1 mol/L KCl. D, CD spectra in 0.1 mol/L KCl. The curves are representative of several experiments. All measurements were performed in a 10-mmol/L lithium cacodylate (pH 7.2) buffer supplemented with 0.1 mol/L potassium chloride. Oligonucleotides were added at 4-μmol/L strand concentration.

Table 1. mRNA expression across the *TOP1* gene 21 exons in the NCI-60

	Mean centered exon intensity									
	Exon									
	1	2	3	4	5	6	7	8	9	10
CO:HCT-116	0.44	0.84	0.97	1.02	0.87	1.14	1.18	1.07	0.84	1.16
BR:MCF7	0.39	0.76	0.96	0.90	0.88	0.96	1.00	0.86	0.68	0.84
LE:CCRF-CEM	0.36	0.63	0.20	0.20	0.50	0.68	0.38	0.24	0.60	0.49
LE:MOLT-4	0.31	0.81	0.58	0.43	0.61	0.63	0.19	0.52	0.46	0.54
LC:NCI-H322M	0.60	1.39	1.13	1.17	1.02	1.12	1.18	1.06	0.96	1.06
CO:HT29	0.17	0.24	0.55	1.04	0.50	0.62	0.98	0.52	0.33	0.61
OV:OVCAR-5	0.25	0.43	0.46	0.46	0.67	0.40	0.46	0.37	0.33	0.48
CO:COLO205	−0.05	0.47	0.34	0.47	0.16	0.37	0.24	0.39	0.26	0.43
LE:K-562	0.21	0.77	0.13	−0.09	0.05	0.00	−0.24	0.11	0.18	0.10
CO:KM12	0.21	0.60	0.25	0.70	0.47	0.63	0.26	0.45	0.42	0.61
PR:PC-3	−0.04	0.63	0.50	0.54	0.04	0.03	0.26	0.16	−0.12	0.18
LC:NCI-H23	0.08	0.41	0.68	0.71	0.43	0.57	0.75	0.57	0.38	0.65
LE:SR	0.22	1.06	1.23	1.49	0.73	0.93	1.27	1.09	0.69	0.99
ME:UACC-62	0.19	0.48	0.49	0.63	0.27	0.57	0.66	0.22	0.37	0.58
CNS:SF-268	−0.15	0.31	0.53	0.49	0.26	0.55	0.79	0.29	0.21	0.54
CO:SW-620	0.00	−0.36	−0.20	0.10	0.02	0.04	−0.03	0.00	0.24	0.19
LC:NCI-H460	0.14	−0.26	−0.06	−0.26	−0.17	0.05	0.18	−0.11	0.05	−0.02
CO:HCT-15	−0.02	0.42	0.49	0.36	0.33	0.55	0.91	0.55	0.30	0.66
ME:LOXIMVI	−0.01	0.08	0.70	1.07	0.29	0.58	1.08	0.70	0.27	0.52
ME:MDA-MB-435	0.12	0.17	0.62	1.07	0.26	0.53	0.43	0.08	0.20	0.23
OV:IGROV1	0.18	−0.44	−0.81	−0.26	−0.21	−0.10	−0.71	−0.58	0.22	−0.35
OV:OVCAR-4	0.09	0.05	−0.40	−0.23	0.19	−0.02	−0.34	−0.14	−0.05	−0.33
BR:MDA-MB-231	−0.05	0.31	0.01	0.19	−0.01	−0.09	0.08	0.09	0.20	0.14
LE:RPMI-8226	0.30	−0.28	0.07	0.02	0.23	−0.32	−0.94	0.19	0.15	0.10
LE:HL-60	−0.03	−0.65	−0.16	−0.05	−0.31	−0.19	−0.30	−0.36	−0.06	−0.27
RE:UO-31	0.02	0.42	0.12	0.40	−0.06	−0.32	−0.01	0.27	−0.09	−0.05
ME:UACC-257	0.00	0.34	0.28	0.38	0.33	0.44	−0.12	0.19	0.13	0.16
OV:OVCAR-3	0.01	0.14	0.11	−0.02	0.18	0.17	0.05	0.08	0.17	0.10
LC:HOP-92	−0.11	−0.63	−0.36	−1.03	−0.29	−0.59	−0.58	−0.25	−0.29	−0.40
CNS:SF-295	−0.02	−0.01	0.12	−0.26	0.01	0.53	0.31	0.13	0.16	0.16
OV:SK-OV-3	0.01	−0.20	−0.49	−0.42	−0.20	−0.12	−0.77	−0.43	−0.31	−0.63
OV:OVCAR-8	−0.04	−0.32	−0.08	−0.44	0.01	0.14	0.10	−0.12	−0.03	−0.24
OV:NCI-ADR-RES	0.06	−0.97	−0.99	−1.55	−0.28	−0.71	−0.99	−0.93	−0.01	−0.59
CO:HCC-2998	−0.10	−0.45	0.08	0.16	0.06	0.34	0.04	0.11	0.12	0.45
ME:MDA-N	−0.04	0.60	0.45	0.44	0.23	0.08	0.61	0.33	0.12	0.36
LC:NCI-H226	−0.18	0.48	0.19	−0.43	−0.11	0.05	0.70	0.38	−0.26	0.01
PR:DU-145	0.07	−0.60	−0.46	−0.56	0.01	−0.28	−0.26	−0.77	−0.49	−0.38
LC:NCI-H522	−0.13	−0.11	−0.65	−0.49	−0.45	−0.72	−0.39	−0.75	−0.22	−0.51
CNS:SNB-19	−0.19	0.16	−0.49	0.14	−0.20	0.11	0.26	0.04	0.12	0.17
BR:T47D	−0.15	−0.80	−0.37	−0.47	−0.17	−0.18	−0.04	−0.55	−0.30	−0.14
LC:HOP-62	−0.12	−0.32	−0.10	−0.42	−0.41	−0.08	0.03	−0.25	0.01	−0.34
BR:BT-549	−0.28	−0.52	−0.37	−0.10	−0.55	−0.69	−0.26	0.21	−0.57	−0.53
ME:SK-MEL-5	−0.10	−0.35	0.02	−0.44	−0.26	−0.31	−0.45	−0.24	−0.18	−0.35
CNS:SF-539	na	na	na	na	na	na	na	na	na	na
RE:TK-10	−0.03	−0.06	−0.14	−0.35	−0.11	−0.18	−0.54	−0.35	−0.30	−0.46
LC:EKVX	−0.09	−0.21	−0.47	−0.48	−0.39	−0.34	−0.61	−0.15	−0.30	−0.19
CNS:U251	−0.12	−1.09	−0.98	−0.29	−0.32	−0.52	−0.91	−0.32	−0.31	−0.44
RE:SN12C	−0.11	0.26	−0.14	−0.20	−0.23	−0.26	−0.03	0.06	−0.39	−0.26
CNS:SNB-75	−0.33	−0.04	−0.32	−0.04	−0.42	−0.20	0.16	−0.21	−0.29	−0.27

(*Continued on the following page*)

Table 1. mRNA expression across the *TOP1* gene 21 exons in the NCI-60 (Cont'd)

Mean centered exon intensity										
Exon										
11	12	13	14	15	16	17	18	19	20	21
0.79	0.93	0.98	0.82	0.93	1.19	0.80	1.07	1.35	0.87	0.84
0.50	0.62	0.83	0.62	0.54	0.81	0.68	0.79	0.91	0.64	0.73
0.43	0.65	0.55	0.67	0.58	0.36	0.72	0.52	0.47	0.50	0.54
0.71	0.46	0.59	0.60	0.48	0.31	0.53	0.48	0.64	0.41	0.46
0.96	1.06	0.69	0.91	1.09	0.53	0.54	0.60	0.46	0.50	0.36
0.19	0.50	0.41	0.35	0.36	0.37	0.44	0.52	0.54	0.39	0.32
0.31	0.65	0.61	0.64	0.62	0.55	0.63	0.67	0.45	0.78	0.75
0.37	0.25	0.12	0.12	0.14	0.42	0.06	0.18	0.39	0.13	0.21
0.04	0.11	−0.04	0.14	0.10	0.12	0.16	−0.01	0.10	0.18	−0.03
0.36	0.50	0.36	0.31	0.36	0.19	0.40	0.43	0.55	0.43	0.19
0.13	−0.06	0.20	−0.01	0.08	0.29	−0.26	0.32	−0.22	−0.13	0.05
−0.31	0.44	0.56	0.43	0.47	0.94	0.30	0.55	0.45	0.44	0.41
0.78	0.67	0.87	0.50	0.63	1.03	0.51	0.81	0.90	0.69	0.67
0.34	0.62	0.59	0.41	0.51	0.63	0.35	0.54	0.29	0.39	0.37
0.14	0.20	0.31	0.04	0.10	0.41	0.00	0.33	0.32	0.23	0.19
−0.16	0.22	0.04	0.14	0.15	−0.02	0.25	0.00	0.38	0.18	−0.03
−0.25	0.03	−0.02	0.02	−0.09	0.11	0.04	0.04	0.10	−0.18	−0.22
−0.13	0.29	0.45	0.30	0.44	0.50	0.38	0.43	0.74	0.44	0.28
0.61	0.27	0.41	0.06	0.22	0.59	0.10	0.40	0.32	0.35	0.24
0.43	0.32	0.30	0.12	0.18	0.16	0.18	0.34	0.20	0.32	0.29
−0.22	0.26	−0.29	0.04	0.04	−0.52	0.33	−0.28	0.03	0.30	−0.10
−0.15	0.05	−0.22	0.20	0.07	−0.18	0.23	−0.06	0.08	0.16	0.02
0.07	−0.09	0.18	0.00	0.07	0.06	−0.08	0.15	0.32	−0.02	0.11
0.58	0.11	0.04	0.26	0.20	−0.06	0.17	0.18	0.12	0.26	0.20
−0.16	−0.10	−0.33	−0.34	−0.27	−0.40	−0.11	−0.25	−0.11	−0.26	−0.19
−0.20	−0.19	0.04	0.06	0.15	0.31	−0.12	0.00	−0.18	−0.02	0.32
0.17	0.41	0.31	0.12	0.35	−0.04	0.25	0.20	0.41	0.34	0.28
0.29	0.25	0.24	0.13	0.16	0.19	0.17	0.15	0.10	0.25	0.15
−0.09	−0.29	−0.06	0.06	−0.08	−0.37	−0.08	−0.24	−0.38	−0.05	−0.26
−0.04	0.07	0.14	−0.03	0.06	0.06	0.08	0.11	−0.21	0.18	−0.07
−0.29	−0.16	−0.20	−0.19	−0.14	−0.36	−0.03	−0.37	−0.23	−0.22	−0.12
0.00	0.00	0.12	0.12	0.16	−0.20	0.06	−0.05	−0.22	0.06	0.11
0.02	0.06	−0.45	−0.09	−0.15	−0.77	0.31	−0.52	−0.19	0.15	−0.16
−0.42	0.11	0.13	−0.01	0.00	−0.17	0.01	0.20	0.32	−0.05	−0.05
0.11	0.02	−0.01	0.25	0.19	0.38	0.03	0.13	0.13	0.26	0.21
−0.37	−0.18	0.07	−0.06	0.03	−0.03	−0.25	−0.27	−0.22	−0.13	−0.17
0.04	−0.13	−0.32	−0.22	−0.14	−0.25	−0.15	−0.46	−0.73	−0.10	−0.29
−0.80	−0.45	−0.50	−0.29	−0.21	−0.48	−0.28	−0.18	0.10	−0.45	−0.34
0.37	−0.07	0.11	0.02	−0.15	0.13	0.07	0.02	−0.05	−0.20	0.00
−0.33	−0.39	−0.42	−0.29	−0.26	−0.57	−0.22	−0.36	−0.29	−0.18	−0.31
−0.36	−0.16	−0.20	−0.24	−0.22	−0.29	−0.06	−0.17	−0.18	−0.12	−0.20
−0.15	−0.46	−0.17	−0.27	−0.35	−0.37	−0.51	−0.29	−0.61	−0.48	−0.41
−0.08	−0.27	−0.34	−0.23	−0.22	−0.27	−0.07	−0.33	−0.26	−0.27	−0.26
na	na	na	na	na	na	na	na	na	na	na
−0.33	−0.38	−0.26	−0.15	−0.29	−0.32	−0.35	−0.32	−0.28	−0.31	−0.18
−0.29	−0.23	−0.27	−0.17	−0.21	0.02	−0.32	−0.29	−0.29	−0.40	−0.45
−0.04	−0.27	−0.51	−0.23	−0.36	−0.61	−0.25	−0.40	−0.21	−0.44	−0.53
−0.24	−0.43	−0.07	−0.28	−0.26	0.07	−0.42	−0.14	−0.30	−0.42	−0.26
−0.57	−0.26	−0.22	−0.34	−0.38	−0.11	−0.35	−0.34	−0.39	−0.39	−0.42

(*Continued on the following page*)

Table 1. mRNA expression across the *TOP1* gene 21 exons in the NCI-60 (Cont'd)

	Mean centered exon intensity									
	Exon									
	1	2	3	4	5	6	7	8	9	10
ME:SK-MEL-2	−0.06	0.34	0.59	0.72	0.20	0.37	1.08	0.27	0.07	0.13
LC:A549	−0.04	−1.32	−0.91	−0.74	−0.71	−1.06	−0.98	−0.80	−0.49	−0.74
ME:M14	−0.18	−0.39	0.10	−0.44	−0.12	−0.18	0.22	−0.09	−0.26	−0.14
ME:MALME-3M	−0.26	−0.19	0.09	−0.02	−0.32	0.01	−0.47	−0.38	−0.24	−0.24
RE:ACHN	−0.06	0.13	−0.66	−1.02	−0.34	−0.91	−1.68	−0.83	−0.57	−0.74
RE:CAKI-1	−0.14	0.30	−0.10	−0.10	−0.17	−0.44	−0.79	−0.06	−0.32	−0.14
RE:A498	−0.10	−0.38	−0.46	−0.54	−0.41	−0.42	−0.23	−0.50	−0.40	−0.41
RE:786-0	−0.20	−0.05	−0.18	−0.20	−0.26	−0.30	−0.25	−0.03	−0.24	−0.04
RE:RXF-393	−0.18	−0.73	−0.44	−0.71	−0.26	−0.48	−0.83	−0.38	−0.36	−0.49
ME:SK-MEL-28	−0.28	−0.04	−0.43	−0.95	−0.52	−0.66	−0.22	−0.45	−0.56	−0.88
BR:HS578T	−0.58	−1.16	−0.99	−1.11	−1.14	−1.04	−0.82	−0.77	−0.90	−0.99
SD	0.20	0.56	0.53	0.63	0.41	0.53	0.65	0.48	0.38	0.50
Mean of intensity	9.21	7.65	7.26	5.98	9.12	8.16	6.55	8.06	9.05	8.15

NOTE: Cell lines ordered as in Fig. 1A.

DNA copy numbers ranged from 4.13 for the breast cancer line MCF-7 to 1.71 for the melanoma SK-MEL-28 (Fig. 2A, left). Comparison of the average transcript z scores (Fig. 1A, last column of the tabular data and Fig. 2A, right) with the average estimated DNA copy number indicates a statistically significant correlation ($r = 0.33$, $P < 0.02$, without multiple comparisons correction). Those results suggest that *TOP1* amplification can contribute to increased expression levels in cancer cells.

Within the portion of chromosome 20 for which copy number was estimated (i.e., from nucleotides 37,103,339 to 41,095,315), 57 of the cell lines showed relative invariance in their average copy number (within the limits of reliability and variability of the individual probes). HS578T exemplifies that type of profile (Fig. 2B, middle). Only two cell lines, breast cancer MCF7 and lung cancer NCI-H322M, showed variation around the *TOP1* locus (Fig. 2B, top and bottom, respectively). In MCF7, there is amplification of an $\sim 1.7 \times 10^6$ nucleotide region that contains TOP1 (Fig. 2B, red squares). In the NCI-H322M cells, there is an ~2-fold increased copy number toward the p-end of chromosome 20 and a reduction in average copy number at the TOP1 locus.

Exon analysis of TOP1 reveals differential expression between the first and the other exons. To expand our transcript assessment and consider potential exon-specific variation, we next examined the intensity levels for each of the 21 *TOP1* exons using the GH Exon 1.0 ST Array (see Supplementary Table S1). This platform was in agreement with the other three transcript platforms for TOP1 expression, with average correlations of 0.86, 0.82, and 0.73 (all significant with $P < 0.0001$) for HG-U95, HG-U133, and HG-U133 Plus 2.0, respectively. Analysis of the GH Exon 1.0 ST Array indicated that the probe intensities for exon 1 were higher on average (by ≥1.87-fold, linear) than those for exons 2 to 21, with the higher expresser cell lines, such as HCT-116 and MCF7 (ranked in Fig. 1A) having less of a drop off (1.37-fold and 1.50-fold linear change, respectively) than the lower expresser cell lines, such as HS578T and SK-MEL-28 (2.23-fold and 2.19-fold linear change, respectively). For all cell lines, the average of the exon 1 minus the average of the exon 2 through 21 probe sets is positive (Supplementary Table S1, last column), indicating the intensity reduction following intron 1.

Next, to partially compensate for variations in probe set hybridization efficiency, we mean-centered the probe set average $\log_2$ intensity values for each exon across the NCI-60 (results in Table 1). Using the data in this form allows one to more clearly identify potential specific exon level fluctuations. Figure 3A shows a clustered image map visualization of these mean-centered intensity values. The vertical strip representing exon 1 stands out as having less color variation and thus less mean-centered intensity variation than do the other exons. The horizontal blocks of color indicate that most cell lines are otherwise consistent in their mean intensity levels across exons 2 to 21. Assessment of the mean-centered $\log_2$ intensities by cell line, including their 95% confidence intervals (Supplementary Fig. S1A, B, and C), indicates that few examples of exon-specific transcript variation occur within exons 2 to 21 for *TOP1*.

The deviations from the (60-cell) mean for the Table 1 data for four individual exons are depicted quantitatively in Fig. 3B, reflecting what was seen in Fig. 3A. That is, from the lengths of the red bars, it is apparent that exon 1 has less deviation from the mean than the other three exons shown. The same is true also for the other 17 exons (for all exons, see Supplementary Fig. S1), substantiating the observation of reduced exon 1 variation from Fig. 3A. That point is reinforced in tabular form as well by the SDs of the mean-centered $\log_2$ intensities for each exon across the NCI-60 (Table 1, second row from bottom). The SD is the least for exon 1, at 0.20, with an average of 0.47 (with a range of 0.35–0.67) for exons 2 to 21.

Table 1. mRNA expression across the *TOP1* gene 21 exons in the NCI-60 (Cont'd)

Mean centered exon intensity										
Exon										
11	12	13	14	15	16	17	18	19	20	21
–0.11	0.09	0.19	–0.07	–0.01	0.28	–0.10	0.16	0.00	0.12	0.04
–0.54	–0.52	–0.74	–0.48	–0.40	–0.65	–0.26	–0.43	–0.21	–0.47	–0.58
0.08	–0.19	–0.19	–0.14	–0.17	–0.29	–0.21	–0.09	–0.18	–0.12	–0.02
–0.24	0.06	–0.32	–0.29	–0.47	–0.31	–0.25	–0.30	–0.11	–0.26	–0.32
–0.29	–0.41	–0.81	–0.38	–0.46	–0.16	–0.63	–0.33	–0.75	–0.29	–0.28
–0.20	–0.60	–0.62	–0.33	–0.32	–0.29	–0.47	–0.31	–0.27	–0.38	–0.16
–0.56	–0.27	–0.39	–0.36	–0.31	–0.43	–0.41	–0.49	–0.52	–0.46	–0.43
–0.36	–0.44	0.07	–0.39	–0.34	–0.21	–0.40	–0.20	–0.04	–0.30	–0.08
0.37	–0.31	–0.35	–0.32	–0.27	–0.27	–0.29	–0.39	–0.43	–0.35	–0.33
–0.33	–0.53	–0.62	–0.52	–0.60	–0.66	–0.50	–0.67	–0.45	–0.44	–0.49
–0.99	–0.48	–0.99	–0.75	–0.78	–0.94	–0.54	–1.01	–0.66	–0.98	–1.07
0.37	0.40	0.44	0.34	0.35	0.44	0.36	0.39	0.44	0.38	0.35
9.78	9.29	8.95	10.50	9.87	7.18	10.66	8.52	8.74	10.31	9.56

Those results indicate that (*a*) *TOP1* mRNA expression is generally less variable for exon 1 than for exons 2 to 21, (*b*) the TOP1 transcript levels generally decrease between the exons 1 and 2 probe sets, and (*c*) there is a greater drop-off in the lower expressers compared with the higher expresser cell lines.

TOP1 intron 1 contains potential quadruplex-forming G-rich sequences. Due to these transcript variations found between *TOP1* exons 1 and 2, we took a detailed look at intron 1 for potential secondary structure-forming motifs (Fig. 4A). The first intron of human *TOP1* is among the smallest in the TOP1 gene, consisting of 330 bp. Examination of the sequence composition of human intron 1 using the quadruplex-forming G-rich sequences (QGRS) Mapper[15] algorithm revealed six and five potential quadruplex-forming G-rich sequences (depicted as the lighter bars in Fig. 4B) within the coding and transcribed strands, respectively. Chimpanzee, mouse, and rat TOP intron 1 (see Fig. 4B) are relatively conserved in both size and number of potential quadruplex-forming G-rich sequences. *Drosophila* is more disparate, with intron 1 containing 1,346 nucleotides and 4 potential quadruplex-forming G-rich sequences in both its coding and transcribed strands (data not shown).

***Evidence for quadruplex formation* in vitro.** Of the QGRS Mapper algorithm identified potential quadruplex sequences on both DNA strands of Intron 1, manual inspection suggested that the most stable (quadruplexes) were located on the transcribed strand. To confirm the propensity of these motifs to form quadruplexes, we synthesized three oligodeoxynucleotides (21, 25, and 30 bases long) that mimicked parts of intron 1 (located in Fig. 4B beneath the human intron 1, with the sequence depicted in Fig. 5A). Standard physicochemical methods were used to reveal quadruplex formation (32). For the 21- and 25-mers, UV melting profiles showed a clear inverted transition at 295 nm, consistent with G4 quadruplex formation (Fig. 5B; ref. 27). The 21-mer was highly stable (T_m of 75°C in K^+). Furthermore, this T_m depends on the nature of the cation ($K^+ > Na^+$; data not shown). The TDS (Fig. 5C) and CD spectra (Fig. 5D) for both sequences were consistent with G4 formation. In contrast, the 30-mer, despite being very G-rich, did not exhibit the same behavior (no transition at 295 nm and different CDS and TDS). Altogether, those results indicate that two of the three sequences form *bona fide* quadruplexes whereas the longest one (30-mer) forms a duplex.

We then analyzed the molecularity of the 21-mer, which shows a fully reversible melting profile. We measured the T_m at different strand concentrations (between 2 and 30 μmol/L). The T_m was concentration independent (data not shown), confirming that the quadruplexes are intramolecular (33). We could not perform a similar experiment with the 25-mer, because its melting was more complex and it showed hysteresis, probably as a result of the formation of multiple quadruplex species of different molecularities.

The melting experiments indicated that at least two sequences in the transcribed strand tend to form quadruplex structures and that the one formed with the 21-mer is extremely thermally stable. Nevertheless, a T_m value is of limited value for assessing the "strength" of a quadruplex under physiologic conditions; a free energy of transition would be more predictive. We therefore performed a van't Hoff analysis of the melting profiles and found a $\Delta G°_{37°C}$ value of –8.6 kcal/mol. That value provides an indication of the high thermodynamic barrier for unfolding of the G4 structure to let the transcription machinery pass through this structure. However, the value of $\Delta G°_{37°C}$ should be interpreted with caution because it relies on a number of assumptions (such as two-state behavior and temperature-independent enthalpy) and extrapolations with respect to the nature of the cellular environment (33).

Discussion

TOP1 is viewed as a housekeeping gene because its homozygous deletion is embryonic lethal (5) and its expression

is restricted across cell cycle stages (34). In this study we used high throughput microarray analyses to elucidate regulatory genetic determinants of *TOP1* transcript expression across the NCI-60 and explored two potential influences on *TOP1* expression: transcriptional pausing and DNA copy number variation. The correlation between TOP1 expression and copy number of 0.33 was higher than prior multigene correlation averages of 0.29 and 0.23 (16), suggesting that *TOP1* copy number alterations in cancers can influence the transcript level.

To compare *TOP1* transcript levels across the NCI-60 cell lines and across our six Affymetrix probe sets (on HG-U95, HG-U133, and HG-U133 Plus 2.0 microarrays), we used average *z* scores.[9] *z* scores facilitate the integration of data from multiple platforms because they are mean-centered and normalized with respect to SD. Average *z* scores for *TOP1* expression provided a single value for each of the NCI-60 lines (see Figs. 1A, B and 2A). Those values were used to rank the cells and for comparison with aCGH (Fig. 2A). Use of *z* scores, as opposed to individual probe set data, for *TOP1* transcript levels yielded generally higher correlations with both TOP1 protein expression (18) and *TOP1* DNA copy number (Fig. 2A). The TOP1 transcript expression at relatively low levels in the renal and at high levels in the leukemia and in six of the seven colon carcinoma lines (Figs. 1A, B and 2A) was consistent with our recent results using an ELISA assay to measure *TOP1* protein levels (18).

Recent studies have shown that transcription can be regulated for many genes after establishment of the preinitiation complex (35, 36). Having found that TOP1 expression was variable from cell line to cell line across the NCI-60, we looked for differential expression across the 21 exons of the *TOP1* gene using our data from the exon-specific Affymetrix GeneChip Human Exon 1.0 ST microarray. An asset in organizing and visualizing the exon array data was our SpliceCenter tool (37), which provided (*a*) automated visualizations of the translated and untranslated regions of the *TOP1* gene; (*b*) the relative exon sizes and locations of introns; and (*c*) the portions of the exons being assessed by the individual probe sets (examples in Fig. 1C). Interpretation of inter–cell line transcript level variation was made possible by the relatively large number of arrays (156) used in this study. As probe set intensities are not directly interpretable with respect to differences among exons because of variability in hybridization efficiency between individual probes, we developed a mean-centering approach and looked for variation across cell lines for each individual exon. Mean-centered intensities, as seen in Fig. 3 for exon 1, can provide indications of exon-specific transcript variations. Analogous transformation of exon array data from genes other than *TOP1* should likewise be useful for detection of exon-specific transcript variation.

Our analyses based on the GH Exon 1.0 ST microarray revealed only very limited variation in exon 1–specific transcript intensities across the NCI-60 compared with differences across cell lines for the other 20 exons (see Fig. 3). The cell lines with the highest *TOP1* expression (such as colon carcinoma HCT-116) seem to allow the transcription process to pass through intron 1 more readily than do the low-*TOP1* expressers (such as breast carcinoma HS578T). In the latter, transcription seems to be impeded within intron 1. Those observations suggested the existence of a negative transcription regulator within intron 1 of the *TOP1* gene in humans. Further analysis showed that intron 1 is relatively short and conserved among vertebrates (see Fig. 4A and B), containing multiple potential quadruplex-forming G-rich sequences. This implies similar potential for the formation of secondary structures in these vertebrates.

Sequences that form quadruplexes have been generally found to regulate transcription within promoter regions. For instance, quadruplexes in the promoters of the HIF1A, KRAS, CMYB, and CMYC genes (38–42) have been found to repress transcription initiation. Our findings for the *TOP1* gene highlight the potential role of quadruplex-forming G-rich sequences for regulating transcription in the body of genes. They are consistent with a recent report showing the preferential enrichment of potential quadruplex sequences in the first introns of a large number of human genes (43). Potential quadruplexes were significantly overrepresented on the nontemplate DNA strands (43). In the case of the *TOP1* gene, potential quadruplex sequences in the first intron were present in both the transcribed (nontemplate) and coding strands, but the sequences with the highest potential were on the transcribed strand. Biochemical and biophysical evidence provided here indicates that the sequences with potential for G-quadruplex formation can actually form such structures, at least *in vitro* (see Fig. 5).

To the best of our knowledge, our study is the first to use exon-specific expression data for recognition of potential intronic secondary structure leading to the restriction of transcript elongation. The approach, practiced across panels of cell lines such as the NCI-60, could be useful for detection of intronic transcription pausing in other genes.

In summary, the present study describes (*a*) the relative levels of *TOP1* within the NCI-60, (*b*) a significant association between *TOP1* expression and DNA copy number, (*c*) the existence of a reduction of *TOP1* transcript level following exon 1, (*d*) the existence of reduced variability of exon 1 (compared with the other 20 exons), (*e*) the presence of multiple potential guanosine quadruplexes within intron 1, and (*f*) the verification of the guanosine quartet formation by two of these by physicochemical methods. This form of exon-specific transcript variation analysis across the NCI-60 allows the recognition of inhibitory elements within the introns of genes. Based on this approach, we suggest, for the first time, the existence of a transcriptional pausing event in the first intron of TOP1 and provide a plausible mechanistic explanation based on the presence of G-quadruplex regions.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Dr. Thomas Gingeras (Cold Spring Harbor Laboratory) and John Ward for providing the microarrays (Affymetrix) and Eric Kaldjian for

overseeing data generation for the Affymetrix GH Exon 1.0 ST microarrays (GeneLogic).

Grant Support

Intramural Program of National Cancer Institute, Center for Cancer Research.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 09/23/2009; revised 12/10/2009; accepted 12/29/2009; published OnlineFirst 03/09/2010.

References

1. Pommier Y. Topoisomerase I inhibitors: camptothecins and beyond. Nat Rev Cancer 2006;6:789–802.
2. Champoux JJ. DNA topoisomerases: structure, function, and mechanism. Annu Rev Biochem 2001;70:369–413.
3. Wang JC. Cellular roles of DNA topoisomerases: a molecular perspective. Nat Rev Mol Cell Biol 2002;3:430–40.
4. Teicher BA. Next generation topoisomerase I inhibitors: rationale and biomarker strategies. Biochem Pharmacol 2008;75:1262–71.
5. Morham S, Kluckman KD, Voulomanos N, Smithies O. Targeted disruption of the mouse topoisomerase I gene by camptothecin selection. Mol Cell Biol 1996;16:6804–9.
6. Miao ZH, Player A, Shankavaram U, et al. Nonclassic functions of human topoisomerase I: genome-wide and pharmacologic analyses. Cancer Res 2007;67:8752–61.
7. Liu LF, Wang JC. Supercoiling of the DNA template during transcription. Proc Natl Acad Sci U S A 1987;84:7024–7.
8. Sordet O, Redon CE, Guirouilh-Barbat J, et al. Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. EMBO Rep 2009.
9. Madden KR, Champoux JJ. Overexpression of human topoisomerase I in baby hamster kidney cells: hypersensitivity of clonal isolates to camptothecin. Cancer Res 1992;52:525–32.
10. Pommier Y, Barcelo JM, Rao VA, et al. Repair of topoisomerase I-mediated DNA damage. Prog Nucleic Acid Res Mol Biol 2006;81:179–229.
11. Pfister T, Reinhold W, Agama K, et al. Topoisomerase I levels in the NCI-60 cancer cell line panel: comparison of enzyme levels determined by a validated ELISA and microarray mRNA levels. Mol Cancer Ther 2009;8:1878–84, Epub 2009 Jul 7.
12. Nitiss J, Wang JC. DNA topoisomerase-targeting antitumor drugs can be studied in yeast. Proc Natl Acad Sci U S A 1988;85:7501–5.
13. Antony S, Agama KK, Miao ZH, et al. Novel indenoisoquinolines NSC 725776 and NSC 724998 produce persistent topoisomerase I cleavage complexes and overcome multidrug resistance. Cancer Res 2007;67:10397–405.
14. Braun MS, Richman SD, Quirke P, et al. Predictive biomarkers of chemotherapy efficacy in colorectal cancer: results from the UK MRC FOCUS trial. J Clin Oncol 2008;26:2690–8.
15. Burgess DJ, Doles J, Zender L, et al. Topoisomerase levels determine chemotherapy response *in vitro* and *in vivo*. Proc Natl Acad Sci U S A 2008;105:9053–8.
16. Bussey KJ, Chin K, Lababidi S, et al. Integrating data on DNA copy number with gene expression levels and drug sensitivities in the NCI-60 cell line panel. Mol Cancer Ther 2006;5:853–67.
17. Scherf U, Ross DT, Waltham M, et al. A gene expression database for the molecular pharmacology of cancer. Nat Genet 2000;24:236–44.
18. Pfister TD, Reinhold WC, Agama K, et al. Topoisomerase I levels in the NCI-60 cancer cell line panel determined by validated ELISA and microarray analysis and correlation with indenoisoquinoline sensitivity. Mol Cancer Ther 2009;8:1878–84.
19. Boyd MR, Paull KD. Some practical considerations and applications of the National Cancer Institute *in vitro* anticancer drug discovery screen. Drug Dev Res 1995;34:91–109.
20. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer 2006;6:813–23.
21. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet 2000;24:227–35.
22. Weinstein JN. Spotlight on molecular profiling: "integromic" analysis of the NCI-60 cancer cell lines. Mol Cancer Ther 2006;5:2601–5.
23. Weinstein JN, Pommier Y. Transcriptomic analysis of the NCI-60 cancer cell lines. C R Biol 2003;326:909–20.
24. Shankavaram UT, Reinhold WC, Nishizuka S, et al. Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. Mol Cancer Ther 2007;6:820–32.
25. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 2003;31:e15.
26. Cantor CR, Warshaw MM, Shapiro H. Oligonucleotide interactions: 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. Biopolymers 1970;9:1059–77.
27. Mergny JL, Phan AT, Lacroix L. Following G-quartet formation by UV-spectroscopy. FEBS Lett 1998;435:74–8.
28. Guédin A, De Cian A, Gros J, Lacroix L, Mergny JL. Sequence effects in single-base loops for quadruplexes. Biochimie 2008;90:686–96.
29. Saccà B, Lacroix L, Mergny JL. The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. Nucleic Acids Res 2005;33:1182–92.
30. Mergny JL, Li J, Lacroix L, Amrane S, Chaires JB. Thermal difference spectra: a specific signature for nucleic acid structures. Nucleic Acids Res 2005;33:e138.
31. Eisenberg E, Levanon EY. Human housekeeping genes are compact. Trends Genet 2003;19:362–5.
32. Mergny JL, Lacroix L. UV melting of G-quadruplexes. Curr Protoc Nucleic Acid Chem 2009;Chapter 17:Unit 17 1.
33. Mergny JL, Lacroix L. Analysis of thermal melting curves. Oligonucleotides 2003;13:515–37.
34. Heck MM, Hittelman WN, Earnshaw WC. Differential expression of DNA topoisomerases I and II during the eukaryotic cell cycle. Proc Natl Acad Sci U S A 1988;85:1086–90.
35. Core LJ, Lis JT. Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 2008;319:1791–2.
36. Margaritis T, Holstege FC. Poised RNA polymerase II gives pause for thought. Cell 2008;133:581–4.
37. Ryan MC, Zeeberg BR, Caplen NJ, et al. SpliceCenter: a suite of web-based bioinformatic applications for evaluating the impact of alternative splicing on RT-PCR, RNAi, microarray, and peptide-based studies. BMC Bioinformatics 2008;9:313.
38. Qin Y, Hurley LH. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. Biochimie 2008;90:1149–71.
39. Cogoi S, Xodo LE. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. Nucleic Acids Res 2006;34:2536–49.
40. De Armond R, Wood S, Sun D, Hurley LH, Ebbinghaus SW. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1α promoter. Biochemistry 2005;44:16341–50.
41. Palumbo SL, Memmott RM, Uribe DJ, Krotova-Khan Y, Hurley LH, Ebbinghaus SW. A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. Nucleic Acids Res 2008;36:1755–69.
42. Belotserkovskii BP, De Silva E, Tornaletti S, Wang G, Vasquez KM, Hanawalt PC. A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. J Biol Chem 2007;282:32433–41.
43. Eddy J, Maizels N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. Nucleic Acids Res 2008;36:1321–33.