

Distortion of quantitative genomic and expression hybridization by C₀t-1 DNA: mitigation of this effect

Heather L. Newkirk, Joan H.M. Knoll and Peter K. Rogan¹

Laboratories of Genomic Disorders and ¹Human Molecular Genetics, Children's Mercy Hospital, University of Missouri-Kansas City School of Medicine, KS, USA

Received October 21, 2005; Revised and Accepted November 23, 2005

ABSTRACT

Cross-hybridization of repetitive sequences in genomic and expression arrays is reported to be suppressed with repeat-blocking nucleic acids (C₀t-1 DNA). Contrary to expectation, we demonstrated that C₀t-1 also enhanced non-specific hybridization between probes and genomic targets. When added to target DNA, C₀t-1 enhanced hybridization (2.2- to 3-fold) to genomic probes containing conserved repetitive elements. In addition to repetitive sequences, C₀t-1 was found to be enriched for linked single copy (sc) sequences. Adventitious association between these sequences and probes distort quantitative measurements of the probes hybridized to desired genomic targets. Quantitative microarray hybridization studies using C₀t-1 DNA are also susceptible to these effects, especially for probes that map to genomic regions containing conserved repetitive sequences. Hybridization measurements with such probes are less reproducible in the presence of C₀t-1 than for probes derived from sc regions or regions containing divergent repeat elements, a finding with significant ramifications for genomic and expression microarray studies. We mitigated the requirement for C₀t-1 either by hybridizing with computationally defined sc probes lacking repeats or by substituting synthetic repetitive elements complementary to sequences in genomic probes.

INTRODUCTION

Genome-wide analysis of gene expression and locus copy number has been facilitated by microarray and array-based comparative genomic hybridization. Persistent questions

regarding reproducibility of these techniques have been raised by cross-validation studies in different laboratories (1–5). Strategies to mitigate variability in the results obtained from replicate studies have focused on standardizing technical factors, such as array production, RNA synthesis, labeling, hybridization, scanning and data analysis (6–8). Zakharkin *et al.* (9) suggest that biological differences among samples are the largest source of this variability and these other factors contribute to a lesser degree.

The use of repetitive sequence-enriched (C₀t-1) DNA to suppress non-specific cross hybridization between repetitive elements present in the probe with other locations in the genome (or transcriptome) is a common requirement for most microarray hybridization studies. In humans, the C₀t-1 fraction is highly concentrated in families of interspersed repetitive elements, such as SINEs and LINEs (10,11). Commercial procedures for C₀t-1 DNA preparation iterate denaturation and re-annealing of genomic DNA, and are monitored by enrichment for Alu elements (3-fold excess over the corresponding level in the normal genome) and L1 elements (4-fold excess). Current quality control procedures do not determine the precise composition of C₀t-1 DNA.

While the C₀t-1 fraction appears to suppress repetitive sequence hybridization, it also increases experimental noise (12). We investigated the possibility that differences in C₀t-1 composition could be a major source of variability in results from genomic hybridization studies. Here, the role of C₀t-1 in genomic hybridization is elucidated by quantitative microsphere hybridization (QMH) (H. L. Newkirk, P. K. Rogan, M. Miralles and J. H. Knoll, manuscript submitted) using sequence-defined, genomic single copy (sc) probes (13) and probes composed of contiguous sc and repetitive genomic sequences. We find that C₀t-1 promotes the formation of stable duplexes containing paralogous repetitive sequences often unrelated to the probe, thereby altering accurate quantification of sc sequence hybridization. We eliminated this effect by developing probes lacking repetitive sequences and by suppressing cross-hybridization with synthetic repetitive elements.

*To whom correspondence should be addressed. Tel: +1 816 983 6511; Fax: +1 816 983 6515; Email: progan@cmh.edu

Table 1. QMH (mean fluorescence)

Reaction	Target DNA ^a	Repeat-blocking agent (ng) ^b	Probe	Geometric mean FL2-SPE ^a	FL1-FITC ^b
Effects of C ₀ t-1 on hybridization intensity levels affected by genomic location of probe					
1	Genomic	0	ABL1a	105.62	N/A
2	Genomic	C ₀ t-1 (50)	ABL1a	235.19	N/A
3	Genomic	0	PMP22	433.76	N/A
4	Genomic	C ₀ t-1 (50)	PMP22	469.27	N/A
5	Genomic	0	TEKT3	642.68	N/A
6	Genomic	C ₀ t-1 (50)	TEKT3	734.04	N/A
7	Genomic	0	HOXB1	890.91	N/A
8	Genomic	C ₀ t-1 (50)	HOXB1	821.35	N/A
9	Genomic	0	HOXB1	332.94	N/A
10	Genomic	C ₀ t-1 (50)	HOXB1	279.1	N/A
11	Genomic	0	HOXB1	2034.76	N/A
12	Genomic	C ₀ t-1 (50)	HOXB1	1727.8	N/A
Dual detection of genomic target and C ₀ t-1 DNA in single reactions					
13	Genomic	0	ABL1a	187.02	5.67
14	Genomic	C ₀ t-1 (50)	ABL1a	390.79	282.48
15	pUC19	C ₀ t-1 (50)	ABL1a	5.88	6.34
Dilution series of C ₀ t-1 DNA in hybridization reactions					
16	Genomic	C ₀ t-1 (50)	ABL1b	304.91	77.8
17	Genomic	C ₀ t-1 (100)	ABL1b	407.61	141.41
18	Genomic	C ₀ t-1 (150)	ABL1b	449.94	234.44
Hybridization experiments with recovered products					
19	Genomic	0	ABL1a	153.12	5.37
20	Genomic	C ₀ t-1 (50)	ABL1a	339.8	191.57
21	Genomic ^c	0	ABL1AluMER1	4.55	3.27
22	Genomic ^c	C ₀ t-1 (50) ^c	ABL1AluMER1	5.36	35.32
Genomic reconstruction experiments					
23	PCR	0	C1QTNF7	270.72	N/A
24	PCR	C1QTNF7LTR (500)	C1QTNF7	270.42	N/A
25	PCR	C ₀ t-1 (50)	C1QTNF7	321.02	N/A
26	Genomic	0	C1QTNF7	1001.83	N/A
27	Genomic	C1QTNF7LTR (500)	C1QTNF7	806.59	N/A
28	Genomic	C ₀ t-1 (50)	C1QTNF7	1226.61	N/A
29	Genomic	0	ABL1a	565.73	N/A
30	Genomic	ABL1aAlu, ABL1aL2 (500)	ABL1a	554.27	N/A
31	Genomic	C ₀ t-1 (50)	ABL1a	1205.01	N/A
32	PCR	0	HOXB1b	94.66	N/A
33	PCR	HOXB1AluL1 (500)	HOXB1b	28.41	N/A

^aNick-translated using biotin-16dUTP and detected using SPE on FL2, 50 ng per reaction.

^bNick-translated using digoxigenin-11dUTP and detected using FITC on FL1.

^cRecovered products from previous hybridization assay.

MATERIALS AND METHODS

Quantitative microsphere hybridization

Probe selection, synthesis and microsphere conjugation. Sc or mixed sc and repetitive sequence probes were designed as described previously (13–15). They include: two chromosome 9q34 probes with sc intervals (i) ABL1a (May 2004, chr 9: 130 623 551–130 625 854) with divergent AluJo/Sx/L2 repeats and (ii) ABL1b (chr 9: 130 627 353–130 628 735) with divergent AluJo repeats from within *ABL1*, designed and validated previously (H. L. Newkirk, P. K. Rogan, M. Miralles and J. H. Knoll, manuscript submitted, 14), (iii) a 1823 bp chromosome 9 probe with Alu/MER1 repetitive sequences, ABL1AluMER1 (chr 9: 130 621 702–130 623 525), (iv) a 98 bp sc segment of a *TEKT3* intron (chr 17: 15 149 108–15 149 206) and (v) a 101 bp sc segment of a *PMP22* intron (chr 17: 15 073 475–15 073 576), (vi) a 93 bp sc segment of a *HOXB1* intron (chr 17: 43 964 237–43 964 330). Probes for genomic reconstruction experiments included: (vii) *HOXB1b* (chr 17: 43 963 396–43 965 681) and (viii) *C1QTNF7* (chr 4: 15 141 452–15 141 500). Repetitive sequences found within

probes were defined as divergent, based on percent sequence differences (>12%), percent deletion (>4%) and/or percent insertion (>4%) relative to consensus family members (www.girinst.org). Probes were synthesized and coupled to microspheres as described previously (H. L. Newkirk, P. K. Rogan, M. Miralles and J. H. Knoll, manuscript submitted).

Genomic target preparation. Genomic template was prepared using four methanol-acetic acid fixed cell pellets derived from cytogenetic preparations of bone marrow samples as described previously (H. L. Newkirk, P. K. Rogan, M. Miralles and J. H. Knoll, manuscript submitted). One microgram of genomic and pUC19 DNA was nick-translated with biotin-16 dUTP to obtain products of 100–350 bp and 50–300 bp in length, respectively (16). One microgram of each C₀t-1 DNA (from Manufacturers I and R) was nick-translated with digoxigenin-11 dUTP to obtain products of 50–300 bp.

Hybridization reactions and flow cytometry. Nick-translated DNA (50 ng) was diluted in 40 µl of 1.5× TMAC hybridization

Table 2. Probes and primers used in this study

Probe	Chromosome	Primer name	Sequence (5'→3')
ABL1a	9	ABL1aF ABL1aR	GTGGCTTATGCCTGTAATTCACA AGAGACAGGGTCTTCTTATGTTGC
ABL1b	9	ABL1bF ABL1bR	ATTTGGAAAGATTATATCCATCTACTTAATGC ACAAACCTACCTACGTTTCAACACTCTCTT
ABL1c	9	ABL1cF ABL1cR	GCTTTATGAAC TAGCTGATTTAGTTTGCTC CTCAATCTCTCTTTTATCTGTTTTGTCCATTG
ABL1d	9	ABL1dF ABL1dR	TAGTTAATTTAGAAGGTTAAATCACGAGAA CTAATTTTTAAATGTGTGAATGCAATTTT
RRP4-1.6a5'	9	RRP4-1.6a5'F RRP4-1.6a5'R	CAGAGGAAGGAAGACGTAGTGAAC GCTGAACCAAGCAGACACAG
RRP4-1.6a	9	RRP4-1.6aF RRP4-1.6aR	ATGGGAGCTTGGATAAGAGATG CTATACCCTGAGGCGATAATGTTC
RRP4-1.6a3'	9	RRP4-1.6a3'F RRP4-1.6a3'R	AGCAGATCAGACATACAGGTCCAA GGCCACCGTAAAGTTACAAGACC
ABL1AluMer1	9	ABL1AluMER1-F ABL1AluMER1-R	C12-Amine-CCTCTCTCGGGGTAGAGTTTCGCTCT CTCAGGCCCTTGTCACACTCTTGAA
DNJA3Alu	16	DNJA3Alu-F DNJA3Alu-R	CTCCTGTCCGTGCTCTGTC AGGCTGGTAGTGACCTGTGG
HOXB1b	17	HOXB1b-F HOXB1b-R	TCACCCCATTCATCTATT TAGGAAGGGGGTAGGGAGTG(-biotin)
HOXB1AluL1	17	HOXB1AluL1-F HOXB1AluL1-R	TCACCCCATTCATCTATT TCCCAAAGTGCTAGGATTGC
C1QTNF7	4	C1QTNF7-F C1QTNF7-R	TGCAATTCAAAACAGATTGAAAAT CCACCATGTGAGAAGTTTGACTAC-biotin
C1QTNF7LTR	4	C1QTNF7LTR-F C1QTNF7LTR-R	AAGTGTGAAAAGGCATATTTTAGCC TACATTTTGGGGTCATTTGTTATG

buffer (3-mol/l tetramethylammonium chloride, 50 mmol/l Tris-HCl, pH8.0, 1 g/l Sarkosyl) containing 10 000 probe-coupled microspheres. Reactions were assembled with the components listed in Table 1. Hybridization and detection of reactions were carried out as described (H. L. Newkirk, P. K. Rogan, M. Miralles and J. H. Knoll, manuscript submitted). In brief, reactions were denatured for 3 min and hybridized overnight at 50°C. Hybridized microspheres were washed and stained with a reporter molecule, streptavidin phycoerythrin (SPE; Molecular Probes, Eugene, OR) and/or anti-digoxigenin-fluorescein isothiocyanate (FITC; Molecular Probes), followed by flow cytometric analysis (FACSCalibur, Becton Dickinson, San Jose, CA). Approximately 5000 microspheres were analyzed per reaction. Hybridization was quantified from the SPE and/or FITC mean fluorescence intensity (measured in channels FL2 and/or FL1, respectively), which corresponds to the quantities of genomic target (FL2) and of C₀t-1 DNA (FL1) bound by probe. Calibration studies with conjugated probes and labeled targets containing identical sequences demonstrated that changes in mean fluorescence intensity were linearly related to the amount of target hybridized. The FL1 and FL2 channel background fluorescence was separately determined in each hybridization experiment using a negative control containing all reaction components except target DNA. Optimal photomultiplier voltages were set as described previously; data collection and analysis were performed with manufacturer-supplied CellQuest software (H. L. Newkirk, P. K. Rogan, M. Miralles and J. H. Knoll, manuscript submitted).

Recovery of probe-hybridized DNA fragments. Aliquots (35 µl) of genomic hybridizations with ABL1a (Table 1, reactions 19 and 20) were washed with 250 µl of 0.1× SSC, 1% SDS and pelleted by centrifugation (13 000 g), and repeated twice. The hybridized genomic sequences

were heat denatured at 95°C for 5 min and snap-cooled followed by centrifugation (13 000 g) at 4°C for 3 min. Recovered sequences were used as target for quantitative PCR (QPCR) and for hybridization to microsphere-coupled ABL1AluMER1 (Table 1, reactions 21 and 22).

Synthetic repetitive DNA. Synthetic repetitive DNA was prepared from genomic regions selected based on the families of repetitive sequences contained within them, since each is enriched in the C₀t-1 manufacturing process. However, any representative genomic region containing sc regions adjacent to moderate or high copy number repetitive elements could have been employed. To demonstrate that repeat elements in genomic probes could be suppressed at locations beyond the desired target interval, we prepared a probe containing a 1.1 kb LTR element centered between two 400 bp sc regions on chromosome 4p (chr 4: 15 139 704–15 141 581) located upstream of the C1QTNF7 gene (Figure 2A). Subsequently, repetitive sequences situated within the ABLa probe region for blocking this repeat element were synthesized; ABL1a contains a 280 bp AluJo repeat, a 300 bp AluSx repeat and an 830 bp L2 element segment (Figure 2C). We also used a 2286 bp segment on chromosome 17q located 5' of HOXB1 containing a 306 bp AluSx repeat and 154 bp L1 truncated sequence (chr 17: 43 963 396–43 965 681) as a probe (Figure 2B). Primers that amplified unique sequences immediately flanking these elements (Table 2, HOXB1AluL1 and C1QTNF7LTR) were developed for PCR amplification of each repeat sequence and of the target product (Table 2, HOXB1b and C1QTNF7). Genomic DNA (Promega) probes were amplified using Pfx (Invitrogen); amplification products were then electrophoresed and extracted by micro-spin column centrifugation. Hybridization reactions (Table 1, reactions 23–33) evaluated the effect of the synthetic repetitive PCR products hybridized to homologous PCR product, and/or

genomic DNA, in the presence and absence of C₀t-1 DNA. Reactions were hybridized, washed, stained with SPE and then analyzed by flow cytometry.

Quantitative PCR

QPCR and data analysis were performed using the Chromo4 QPCR system (BioRad Laboratories, Hercules, CA). Primers and amplified intervals were verified for unique genomic representation using BLAT (17) and WU-BLAST (Table 2). Each 50 μ l reaction contained 0.5 μ M of each primer, 50 ng C₀t-1 template or positive control human genomic DNA (Promega, Madison, WI) and 25 μ l 2XQTSybrG master mix (Qiagen, Valencia, CA). Genomic DNA was nicked using DNase to generate fragments from 50 to 300 bp, and a negative control contained all reaction components except for DNA. Thermal cycling conditions were 95°C for 15 min, 45 cycles of amplification (94°C for 15 s, 61°C for 30 s (data acquisition), 72°C for 30 s), followed by 72°C for 5 s with a decrease in temperature by 20°C every second for the generation of a melt curve. A calibration curve used to determine the amount of input target sequence in the recovered genomic template was generated by varying the amounts of normal genomic template (1, 2, 4, 10 and 20 ng) and by determining the C_T values for each reaction.

The composition of sequences recovered from the ABL1a product hybridization (1 μ l; Table 1, reactions 21 and 22) was determined by QPCR. Primer sets were utilized that amplified several sequences from within the ABL1 region, which were not necessarily homologous to this probe, including ABL1a and ABL1c (chr 9: 130 709 665–130 711 469), ABL1d (chr 9: 130 699 324–130 700 596) as well as primers specific for other unlinked genomic regions such as DNJA3Alu (chr 16: 4 421 138–4 421 200), containing an Alu repeat located 5' of the *DNAJ3* gene, *TEKT3* and *HOXB1* (Table 2). Reactions were performed as described above. A positive control (human genomic DNA) was run for each primer set to represent the initial quantity of genomic DNA originally added to QMH reactions (50 ng). Molar ratios of target sequences recovered from QMH were determined from the quantity of initial template in test samples (interpolated from its C_T value cross-referenced against the standard calibration curve) in the presence and absence of C₀t-1 DNA.

RESULTS

Quantitative microsphere hybridization with C₀t-1 DNA

A FISH-validated, mixed sc and repetitive sequence probe, ABL1a, from the 5' end of IVS1b of the *ABL1* gene containing divergent AluJo/Sx/L2 repeats (chr 9: 130 623 551–130 625 854) was hybridized with nick-translated genomic DNA (H. L. Newkirk, P. K. Rogan, M. Miralles and J. H. Knoll, manuscript submitted). Although we had expected that commercially prepared C₀t-1 DNA would suppress repetitive sequence hybridization, in replicate hybridizations of ABL1a with nick-translated genomic DNA, the mean fluorescence (or hybridization) intensity of labeled genomic target was consistently and significantly increased by 2.2-fold when C₀t-1 was included in replicate hybridizations of ABL1a with nick-translated genomic DNA (Table 1, reactions 1 and 2).

Sc probes derived from chromosome 17 genes, *PMP22* (chr 17: 15 073 475–15 073 576) and *TEKT3* (chr 17: 15 149 108–15 149 206), showed smaller but reproducible increases of 1.08- and 1.14-fold in hybridization intensity in the presence of C₀t-1 DNA (Table 1, reactions 3–6). These experiments suggested that the effects due to C₀t-1 are related to the composition of repetitive sequences surrounding these sc intervals. An sc probe from *HOXB1* (chr 17: 43 964 237–43 964 330) consistently exhibited a small decrease in hybridization intensity with addition of C₀t-1 DNA (Table 1, reactions 7–12) with a 0.84- to 0.92-fold decrease in hybridization intensity for genomic samples tested. The *HOXB1* interval is practically devoid of repetitive sequences (UCSC Genome Browser, May 2004 assembly; <http://genome.ucsc.edu>). The region circumscribing ABL1a contains highly dense, conserved and abundant interspersed SINE (AluJo, AluSx) and less conserved LINE (L2) elements. The *TEKT3* and *PMP22* intervals contain shorter, less abundant and more divergent classes of repeat elements (MIR, MER and L2).

The degree to which addition of C₀t-1 DNA altered target hybridization to the ABL1a probe was determined by comparing hybridizations of biotin-labeled target DNA (detected with streptavidin-phycoerythrin [SPE] in the FL2 channel), a biotin-labeled negative control target (pUC19 plasmid) and each of these with digoxigenin-labeled C₀t-1 DNA (detected by FITC-conjugated anti-digoxigenin in the FL1 channel). The presence of C₀t-1 resulted in a 2-fold increase in the mean fluorescence intensity for ABL1a hybridized to biotin-labeled homologous genomic target sequence. However, the amount of labeled C₀t-1 sequence bound substantially exceeded that necessary for suppression of repetitive sequences in ABL1a, based on a 50-fold increase in intensity relative to reactions in which C₀t-1 sequences were omitted (Table 1, reactions 13 and 14). C₀t-1 binding appears sequence-specific, since hybridization of ABL1a to pUC19 exhibited background level signals (<10¹) regardless of whether C₀t-1 was present (Table 1, reaction 15). These findings suggest that homologous sequences in C₀t-1 are directly binding to the ABL1a probe. Because the ABL1a sequence presumably represents only a small proportion of the C₀t-1 target, it alone cannot account for the increase in observed hybridization.

To determine if the increased signal was related to the quantity of C₀t-1 DNA, varying amounts of digoxigenin-labeled C₀t-1 DNA added to a fixed quantity (50 ng) of biotin-labeled genomic target were hybridized to ABL1b, which is a mixed sc and repetitive probe. ABL1b contains two divergent AluJo repetitive sequences (chr 9: 130 627 353–130 628 735). By doubling the amount of C₀t-1 from 50 to 100 ng in the reaction, probe hybridization to C₀t-1 increased by 1.8-fold and to homologous target by 1.3-fold (Table 1, reactions 16 and 17). Similarly, 150 ng of labeled C₀t-1 DNA increased hybridization to ABL1b by 3-fold over 50 ng C₀t-1, and by 1.5-fold to target DNA (Table 1, reactions 17 and 18). Even though the stoichiometric addition of C₀t-1 DNA dilutes the homologous biotinylated target between 2- and 4-fold, the corresponding hybridization intensity is unexpectedly increased 1.5-fold.

The correlation of C₀t-1 concentration with hybridization intensity suggested that this reaction component promoted the formation of duplex structures containing other sequences

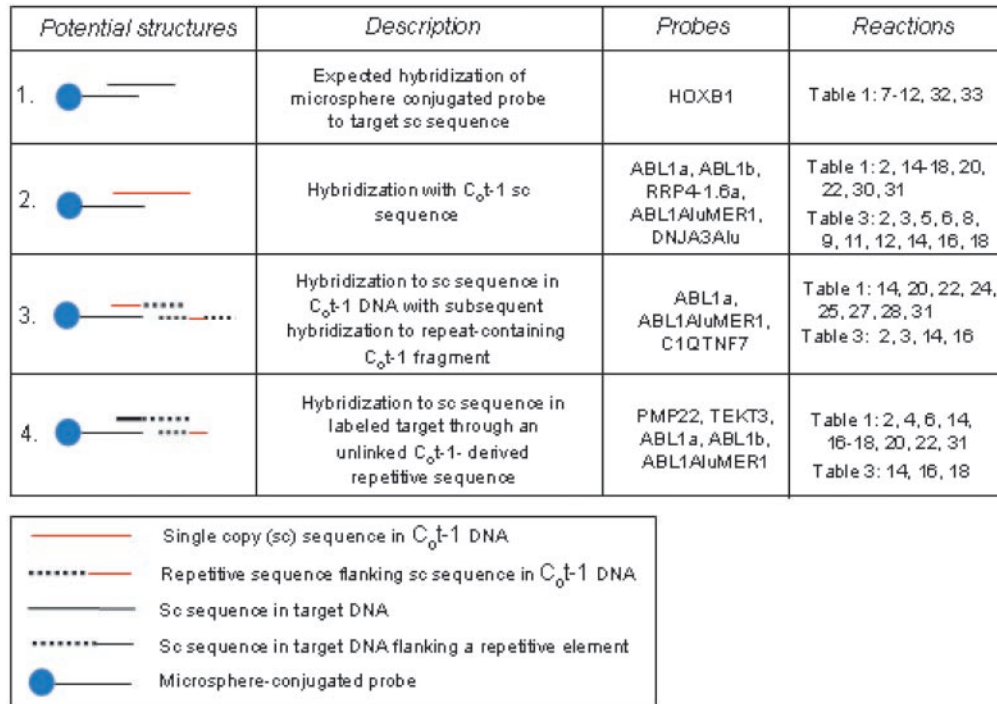


Figure 1. Potential structures produced in QMH hybridization in the presence of C₀t-1 DNA. Example 1 depicts the hybridization of a microsphere-conjugated probe hybridizing to target genomic sc sequence. Example 2 illustrates the hybridization of the probe to sc sequences contained within the C₀t-1 DNA fraction. Example 3 indicates a duplex formed between a sc sequence derived from C₀t-1 DNA which bridged hybridization to a repetitive sequence within C₀t-1 DNA. Example 4 shows a similar event in which the partial duplexes are bridged by unlinked sc sequences within the target genome.

besides the probe and desired genomic target. To determine the composition of C₀t-1 derived sequences bound to probes, products were denatured and recovered after hybridization to ABL1a-coupled microspheres (Table 1, reactions 19 and 20). These products were used as target sequences in subsequent hybridizations to a non-overlapping sc and repetitive microsphere-conjugated probe, ABL1AluMER1, containing Alu elements (AluJb, AluSq, Charlie1 and AluSx) and MER1 sequences localized 2.3 kb centromeric to ABL1a. Given the genomic location of ABL1AluMER1, we did not expect it to be present in the recovered nick-translated genomic products. However, the labeled C₀t-1 fraction was found to be the source of the recovered ABL1AluMER1 sequence, based on a 11-fold increase in mean fluorescence intensity in the FL1 channel (Table 1, reactions 21 and 22). Repetitive sequences adjacent to hybridized ABL1a in C₀t-1 DNA appear to nucleate hybridization to genomic sequences by forming networks of repetitive and sc sequence elements (Figure 1, examples 1 and 2). We evaluated this possibility by QPCR analysis of sequences present in recovered hybridization products.

Analysis of hybridized sequences by QPCR

We determined the content of sc sequences in C₀t-1 that were homologous to our probes by QPCR. A 100 bp sc segment of ABL1a was amplified from 500 ng samples of C₀t-1 DNA and control genomic DNA (Table 2). Based on their respective C_T values, the C₀t-1 fractions from Manufacturers I and R exhibited a 14- and 2-fold increase, respectively, in the amount of ABL1a hybridized (or a 2.5 and 1.7 molar increase) relative to

Table 3. Quantification of recovered hybridization targets

Reaction	Template	Primer set	C _T	ng
1	Genomic DNA (500 ng)	ABL1a	10.88	1.73
2	C ₀ t-1 DNA (Manufacturer I) (500 ng)	ABL1a	7.06	4.25
3	C ₀ t-1 DNA (Manufacturer R) (500 ng)	ABL1a	9.86	3.02
4	Genomic DNA	RRP4-1.6a5'	1.73	N/A
5	C ₀ t-1 DNA (Manufacturer I)	RRP4-1.6a5'	2.22	N/A
6	C ₀ t-1 DNA (Manufacturer R)	RRP4-1.6a5'	1.96	N/A
7	Genomic DNA	RRP4-1.6a	2.00	N/A
8	C ₀ t-1 DNA (Manufacturer I)	RRP4-1.6a	2.75	N/A
9	C ₀ t-1 DNA (Manufacturer R)	RRP4-1.6a	6.63	N/A
10	Genomic DNA	RRP4-1.6a3'	1.88	N/A
11	C ₀ t-1 DNA (Manufacturer I)	RRP4-1.6a3'	1.93	N/A
12	C ₀ t-1 DNA (Manufacturer R)	RRP4-1.6a3'	4.64	N/A
13	Recovered ABL1a hybridized to genomic DNA	ABL1a	9.23	N/A
14	Recovered ABL1a hybridized to genomic DNA with C ₀ t-1	ABL1a	1.36	N/A
15	Recovered ABL1a hybridized to genomic DNA	ABL1AluMER1	24.29	0.009526
16	Recovered ABL1a hybridized to genomic DNA with C ₀ t-1	ABL1AluMER1	18.82	1.38
17	Recovered ABL1a hybridized to genomic DNA	DNJA3Alu	ND	ND
18	Recovered ABL1a hybridized to genomic DNA with C ₀ t-1	DNJA3Alu	30.26	2.053

N/A: not applicable; ND: not detectable.

its normal genomic composition (Table 3, reactions 1-3). ABL1a sequences were recovered after hybridization to determine the composition of genomic and C₀t-1 derived sequences hybridized to this probe (Table 1, reactions 21

and 22). ABL1a sequences were increased by 128-fold in a hybridized sample containing both target and C₀t-1 DNA (Table 3, reactions 13 and 14). Recovered sequences identical to ABL1AluMER1 from hybridizations containing C₀t-1 were 139-fold more abundant than that found in duplicate reactions lacking C₀t-1 (Table 3, reactions 15 and 16). We also detected repetitive sequences that are closely related to ABL1AluMER1 in recovered hybridization products. An Alu element with 92% similarity, DNJA3Alu (5' to DNJA3 gene; chr 16: 4 421 138–4 421 200), was found in the hybridization reaction containing C₀t-1, but not in the reaction lacking C₀t-1, indicating that C₀t-1 was the source of this contaminating sequence (Table 3, reactions 17 and 18). Other sc genomic segments [i.e. from *CMT1A*, *HOXB1* and other *ABL1* regions (ABL1c and ABL1d)] were not detected in the products recovered from hybridization to the ABL1a probe.

C₀t-1 derived sequences hybridized to RRP4-1.6a, a sequence linked to *ABL1* (Table 2), contained both homologous sc and repetitive sequences, despite the fact that this sc probe had been validated by FISH (13). Moderately and highly abundant MIR, L2 and L1 repeat elements surround this sequence in the genome. QPCR demonstrated higher concentrations of repetitive sequences recovered from upstream (5') and downstream (3') amplicons relative to a short RRP4-1.6a product derived from within the sc interval (Table 3, reactions 4–12). Comparison of C_T values indicates that sc sequences bordering genomic repeats (RRP4-1.6a5' and RRP4-1.6a3') are only 6.8-fold more abundant in genomic DNA than in the C₀t-1 fraction for Manufacturer R (and similarly, for Manufacturer I). As expected, the internal sc RRP4-1.6a sequence is considerably more abundant in genomic DNA than in C₀t-1 (24-fold), but nevertheless can still be detected in C₀t-1 (Table 3, reactions 7–9). Enrichment for SINEs and LINEs during C₀t-1 preparation results in accretion of linked sc sequences, which during hybridization can potentially

anneal to the conjugated probe or to actual sc target sequences in labeled genomic DNA.

Suppression of cross-hybridization with synthetic repetitive DNA

We reversed the hybridization effect of C₀t-1 DNA at three different genomic loci by substituting an excess of purified, synthetic DNA(s) prepared specifically from the repetitive elements adjacent to sc sequences (Figure 2). A 1.9 kb amplification product was synthesized containing a LTR-like repetitive element and a sc sequence upstream of *CIQTNF7* on chromosome 4. The addition of the purified synthetic LTR-like element, C1QTNF7LTR, had no effect on the self-hybridization of this product to coupled microspheres, whereas the addition of C₀t-1 DNA increased the mean fluorescence by 1.2-fold (Table 1, reactions 23–25). We also used C1QTNF7LTR to block hybridization of repetitive sequences in nick-translated genomic DNA in the presence and absence of C₀t-1 DNA and obtained similar results (Table 1, reactions 26–28). Hybridization of AluSx and L1 repetitive sequences was suppressed within an ~2.3 kb region on chromosome 17 upstream of the *HOXB1* locus (*HOXB1b*) using a synthetic PCR product, HOXB1AluL1, containing these sequences. Hybridization of the *HOXB1b* PCR product and corresponding microsphere-coupled probe in the presence of the HOXB1AluL1 effectively blocked repetitive sequence within amplified target, and, in fact, reduced hybridization intensity by 0.3-fold, presumably because of the reduction in target length (Table 1, reactions 32 and 33). Hybridization of repetitive sequences was also effectively suppressed in comparable genomic hybridizations to ABL1a coupled to microspheres by addition of synthetic Alu and L2 elements from within this target region (Table 1, reactions 29–31).

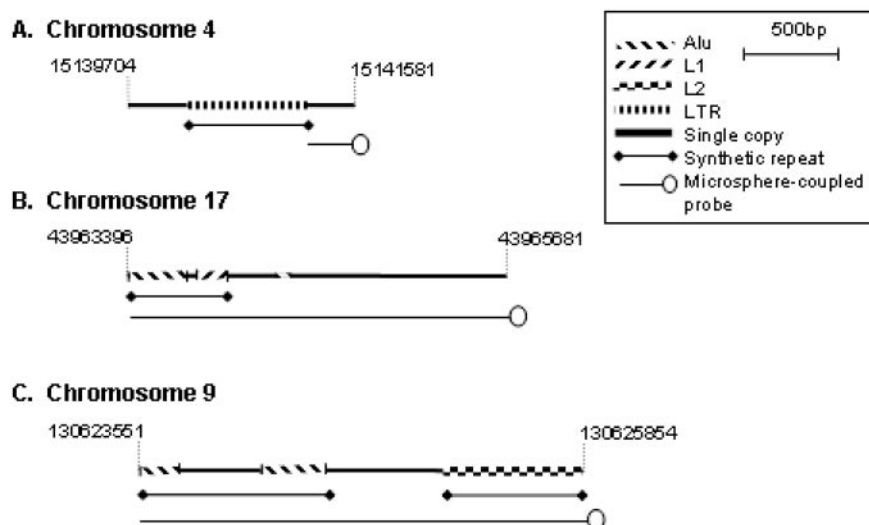


Figure 2. Synthetic repetitive products and probes used in suppression of cross-hybridization to genomic templates. (A) Synthetic repetitive sequence from a 1.1 kb LTR element between two 400 bp sc intervals upstream of the *CIQTNF7* gene on chromosome 4p. The genomic target region for the sc microsphere-coupled probe is downstream of the LTR element. (B) A single synthetic repetitive DNA product contained both the AluSx (306 bp) and L1 (154 bp) repetitive sequences within a 2286 bp segment on chromosome 17q upstream of *HOXB1* (*HOXB1b*). The probe coupled to the microsphere included both sc and repetitive sequences. (C) Individual synthetic DNA products derived from ABL1a contained a 280 bp AluJo repeat, a 300 bp AluSx repeat and an 830 bp L2 element segment. The microsphere-coupled probe contained the entire 2303 bp region.

Table 4. Variation among replicate microarray studies for different genes

Coordinates (hg17)	Gene	Repetitive element(s)	Log ratio range	Difference ^a	Mean	GDS record
Regions containing sc sequences						
Chr17: 43 964 237–43 964 330	<i>HOXB1</i>	N/A	2.44 to 2.51	0.07	2.472	GDS223/31555
Chr17: 15 149 108–15 149 206	<i>TEKT3</i>	N/A	0.15 to 0.21	0.03	0.172	GDS226/729
Chr6: 31 479 350–31 491 069	<i>MICA</i>	N/A	2.91 to 2.93	0.02	2.92	GDS223/40755
Chr6: 33 166 877–33 173 425	<i>HLA-SX-alpha</i>	N/A	2.50 to 2.54	0.04	2.519	GDS223/34072
Chr6: 31 546 991–31 548 163	<i>MHC Class I 3.8-1</i>	N/A	2.43 to 2.44	0.01	2.438	GDS223/34934
Chr6: 31 651 329–31 654 091	<i>TNFalpha</i>	N/A	2.58 to 2.65	0.07	2.614	GDS223/35402
Chr12: 6 179 924–6 217 679	<i>CD9</i>	N/A	2.96 to 2.97	0.01	2.966	GDS223/39389
Regions containing repetitive sequences						
Chr9: 130 623 551–130 625 854	<i>ABL1 (5' IVS1b)</i>	Alu, L2	-0.53 to 0.14	0.67	-0.0975	GDS751/20213
Chr19: 61 581 409–61 581 526	N/A	L1	-0.22 to 1.34	1.56	0.225	GDS221/H200016688
Chr14: 63 636 408–63 636 790	<i>ZFYVE26</i>	L1	-0.25 to 0.53	0.78	0.0825	GDS221/H200018057
Chr6: 43 004 902–43 014 978	<i>TNRC5</i>	Alu, MIR, L1, L2, LTR	-0.25 to 0.45	0.68	0.215	GDS226/11917
ChrX: 129 482 873–129 483 034	N/A	Retrotransposon L1	-0.47 to 0.06	0.53	-0.2225	GDS221/H200014041
Chr14: 67 298 146–67 298 238	N/A	L1, L2	-0.3 to 0.13	0.43	-0.1338	GDS221/H200014930
Chr10: 15 119 774–15 314 575	<i>GAPDH</i>	L2, LTR	1.51 to 2.12	0.61	1.7875	GDS221/H200007830
Chr15: 38 115 535–38 118 631	<i>SRP14</i>	Simple repeat	-0.38 to 0.44	0.82	-0.045	GDS222/H007542

N/A: not assigned.

^aMaximum difference in log ratios.

Impact of C₀t-1 in microarray hybridization

The nearly universal inclusion of C₀t-1 for repeat sequence suppression in published hybridization studies raises the question of how this reagent affects quantitative measures of expression and/or genomic copy number. We evaluated the variability in dual-label hybridization intensities across a set of replicate target samples hybridized to arrays of cloned probes in expression studies which utilized C₀t-1 (source data from the GEO database: <http://www.ncbi.nlm.nih.gov/projects/geo>). We first analyzed results from cDNA probes of genes used in our microsphere hybridization assay (including *ABL1a*, *HOXB1* and *TEKT3*), then subsequently the hybridization profiles for several gene sequences located in genomic environments distinguished by their repetitive sequence composition, i.e. that were either densely (Table 4, bottom) or sparsely (Table 4, top) populated with repetitive sequences. Replicate Cy3/Cy5 intensity ratios are significantly more variable for sequences occurring within repeat-dense genomic intervals relative to probes derived from genomic regions containing fewer, more divergent repetitive sequences. For example, *ABL1* was found to exhibit both increased and decreased expression using the same test sample in different replicates (e.g. Database record GDS751/20213 which displays a sample variance of 0.30, corresponding to $P = 0.18$; Table 4), analogous to the distortion in hybridization we observed with microsphere conjugated-*ABL1a*. In contrast, *HOXB1* showed little variability in log ratio intensities among replicate expression array studies using the same test sample (GDS223/31555; sample variance = 0.001 corresponding to $P < 0.0001$), consistent with our results for this locus. This suggests that sc sequences in C₀t-1 hybridize to probes, nucleating the formation of mixed sc and repetitive sequence networks that capture labeled repetitive sequences from target cDNA. In microarray studies, C₀t-1 thus distorts the hybridization of cloned probes enriched for interspersed repetitive sequences by forming complex hybridization networks in a manner analogous to what is observed in QMH.

DISCUSSION

We have demonstrated that sequences present in C₀t-1 DNA can significantly alter the amount of labeled genomic target detected in hybridization reactions with homologous probes. Rather than suppressing cross-hybridization, C₀t-1 enhanced hybridization to probes containing repetitive sequences by as much as 3-fold. Our results suggest that unlabeled C₀t-1 DNA sequences bridge sc and repetitive sequences in sequence specific probes and complementary target sequences. Repetitive sequences linked to homologous sc sequences in the C₀t-1 fraction can nucleate subsequent hybridization of labeled repetitive sequences in genomic targets. The addition of C₀t-1 DNA to probe hybridizations with labeled genomic templates catalyzes the formation of a network of heteroduplexes homologous to the probe and elsewhere in the genome (Figure 1, example 2). 'Partial' duplexes containing both sc and repetitive sequences (Figure 1, example 3) are facilitated by the addition of C₀t-1 DNA through labeled sc genomic targets to linked repetitive elements (Figure 1, example 4). Labeled repetitive sequences linked to sc genomic target DNA sequences can also alter hybridization intensities, but not to the same extent that C₀t-1 does, owing to its enrichment for both sc and interspersed repetitive sequences.

Since the advent of microarray and array CGH technologies, many researchers have noted concerns about experimental reproducibility (4,5). Perhaps the largest source of variation in relation to cross-hybridization stems from repetitive sequences (7). However, many researchers believe this issue is addressed by blocking repetitive elements with C₀t-1 DNA prior to hybridizing cDNA to an array (6,7). Dong *et al.* (18) found 'some regions of non-repetitive sequences were sufficiently homologous to repetitive sequences to hybridize to the human C₀t-1 DNA fraction' and proposed that this was responsible for skewing hybridization intensities in their microarray results. C₀t-1 affects the reproducibility of hybridization assays by promoting the formation of repetitive sequence bridges between probes and unrelated, labeled genomic targets. It also contains sc sequences that compete with labeled targets

for probe sites. A more extensive genome-wide analysis is warranted to identify other genomic regions that are more likely to be susceptible to this source of systematic error.

The repetitive component in C₀t-1 DNA is fractionated based on re-association kinetics rather than being explicitly defined based on sequence composition. Because it is not contaminated with sc sequences, sequence-defined synthetic repetitive DNA is more effective at blocking cross-hybridization by repetitive sequences in probes to paralogous repetitive genomic targets. Another advantage of a locus-specific synthetic reagent is that divergent repetitive sequences or repeat families that are underrepresented in the C₀t-1 DNA fraction can be synthesized, thereby providing a more accurate and comprehensive repertoire of genomic repeat sequences free of sc sequences.

Nevertheless, replacement of C₀t-1 with a synthetic repetitive DNA reagent (patent pending) that includes all known repetitive elements throughout the genome is probably precluded based on the cost and logistical challenges inherent in its preparation. We suggest that further processing of the C₀t-1 fraction may provide a means of significantly reducing the proportion of contaminating sc sequences. Since, the re-annealed sequences in C₀t-1 are predominantly repetitive in nature, these sequences will co-purify with linked single-stranded sequences which are comprised of sc and non-overlapping repetitive components, adjacent to the re-annealed repetitive sequences in the genome. Treatment of these mixed duplex and single-stranded structures with an obligatory processive exonuclease (i.e. such as Mung Bean nuclease or exonuclease I) should trim single-stranded sequences protruding from duplex DNA. These enzymes should not cleave at mismatched nucleotides (which are common among related members of the same repetitive sequence family) within single-stranded bubbles or within nicked duplexes; however, they will digest single-stranded repetitive and/or sc sequences. This will particularly impact the representation of repetitive elements that commonly show 5' or 3' genomic truncation (i.e. such as observed in L1 retrotransposons) (11). Loss of these sequences could be mitigated by addition of the corresponding synthetic DNA reagents. However, it should be noted that this treatment of C₀t-1 DNA will not eliminate all sc sequences as some of the sc sequences may have re-annealed. Re-annealing of sc sequences, however, would not be kinetically favored.

We suggest that substitution of a partial or completely synthetic blocking reagent composed of defined repetitive sequences in place of C₀t-1 DNA could improve the reproducibility of expression microarray and array comparative genomic hybridizations. This should ultimately lead to standardization of experimental conditions in these widely used procedures. Sc probe technology itself could circumvent the requirement for suppression of repetitive sequence cross-hybridization by selecting sc probes that maximize the genomic separation between neighboring repetitive elements.

NOTE ADDED IN PROOF

A preliminary study demonstrated that cross-hybridization of repetitive sequences is significantly diminished by addition of Mung Bean nuclease-treated C₀t-1 DNA to labeled genomic DNA, as proposed. Results were confirmed by comparing C_T

values of single copy sequences in nuclease treated and untreated C₀t-1 DNA.

ACKNOWLEDGEMENTS

We thank Drs Christopher Gocke and Thomas Schneider for their valuable comments and Dr David Zwick, Ruth Morgan and Lisa Wheeler of the Children's Mercy Hospital Flow Cytometry laboratory for their expert advice. Research was supported in part by 1R21CA95167-02 and by the Katherine B. Richardson Trust (Grant #4186). Funding to pay the Open Access publication charges for this article was provided by the Katherine B. Richardson Trust.

Conflict of interest statement. None declared.

REFERENCES

- Oostlander, A., Meijer, G. and Ylstra, B. (2004) Microarray-based comparative genomic hybridization and its applications in human genetics. *Clin. Genet.*, **66**, 488–495.
- van Hijum, S.A., de Jong, A., Baerends, R.J., Karsens, H.A., Kramer, N.E., Larsen, R., den Hengst, C.D., Albers, C.J., Kok, J. and Kuipers, O.P. (2005) A generally applicable validation scheme for the assessment of factors involved in reproducibility and quality of DNA-microarray data. *BMC Genomics*, **6**, 77.
- Bammler, T., Beyer, R.P., Bhattacharya, S., Boorman, G.A., Boyles, A., Bradford, B.U., Bumgarner, R.E., Bushel, P.R., Chaturvedi, K., Choi, D. et al. (2005) Standardizing global gene expression analysis between laboratories and across platforms. *Nature Methods*, **2**, 351–356.
- Sherlock, G. (2005) Of FISH and chips. *Nature Methods*, **2**, 329–330.
- Dobbin, K.K., Beer, D.G., Meyerson, M., Yeatman, T.J., Gerald, W.L., Jacobson, J.W., Conley, B., Buetow, K.H., Heiskanen, M., Simon, R.M. et al. (2005) Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. *Clin. Cancer Res.*, **11**, 565–572.
- Li, X., Weikuan, G., Mohan, S. and Baylink, D. (2002) DNA microarrays: their use and misuse. *Microcirculation*, **9**, 13–22.
- Wren, J., Kulkarni, A., Joslin, J., Butow, R. and Garner, H. (2002) Cross-hybridization on PCR-spotted microarrays. *IEEE Eng. Med. Biol.*, 71–75.
- Marshall, E. (2004) Getting the noise out of gene arrays. *Science*, **306**, 630–631.
- Zakharkin, S., Kim, K., Mehta, T., Chen, L., Barnes, S., Scheirer, K., Parrish, R., Allison, D. and Page, G. (2005) Sources of variation in Affymetrix microarray experiments. *BMC Bioinformatics*, **6**, 214–225.
- Britten, R. and Kohne, D. (1968) Repeated sequences in DNA. *Science*, **161**, 529–540.
- Rogan, P.K., Pan, J. and Weissman, S.M. (1987) L1 repeat elements in the human epsilon-G gamma globin gene intergenic region: sequence analysis and concerted evolution within this family. *Mol. Biol. Evol.*, **4**, 327–42.
- Carter, N., Fiegler, H. and Piper, J. (2002) Comparative analysis of comparative genomic hybridization microarray technologies: report of a workshop sponsored by the Wellcome Trust. *Cytometry*, **49**, 43–48.
- Rogan, P., Cazcarro, P. and Knoll, J. (2001) Sequence-based design of single-copy genomic DNA probes for fluorescence in situ hybridization. *Genome Res.*, **11**, 1086–1094.
- Rogan, P. and Knoll, J. (2003) Sequence-based in situ detection of chromosomal abnormalities at high resolution. *Am. J. Med. Genet.*, **121**, 245–257.
- Knoll, J.H.M. and Rogan, P.K. (2004) US Patent 6,828,097.
- Knoll, J. and Lichter, P. (2005) *In situ* hybridization to metaphase chromosomes and interphase nuclei. In Dracopoli, N., Haines, J., Korf, B., Moir, D., Morton, C., Seidman, C., Seidman, J. and Smith, D. (eds), *Current Protocols in Human Genetics*. Green-Wiley, NY, Vol. 1, Unit 4.3.
- Kent, W. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
- Dong, S., Wang, E., Hsie, L., Cao, Y., Chen, X. and Gingeras, T.R. (2001) Flexible use of high-density oligonucleotide arrays for single-nucleotide polymorphism discovery and validation. *Genome Res.*, **11**, 1418–1424.