

Methodology article

Open Access

## Methods for high throughput validation of amplified fragment pools of BAC DNA for constructing high resolution CGH arrays

Spencer K Watson\*, Ronald J deLeeuw, Adrian S Ishkanian, Chad A Malloff and Wan L Lam

Address: Department of Cancer Genetics and Developmental Biology. BC Cancer Agency 601 W.10th Ave. Vancouver B.C. V5Z 1L3 Canada

Email: Spencer K Watson\* - [swatson@bccrc.ca](mailto:swatson@bccrc.ca); Ronald J deLeeuw - [rdeleeuw@bccrc.ca](mailto:rdeleeuw@bccrc.ca); Adrian S Ishkanian - [aishkani@bccrc.ca](mailto:aishkani@bccrc.ca); Chad A Malloff - [cmalloff@bccrc.ca](mailto:cmalloff@bccrc.ca); Wan L Lam - [wlam@bccrc.ca](mailto:wlam@bccrc.ca)

\* Corresponding author

Published: 14 January 2004

Received: 17 October 2003

BMC Genomics 2004, 5:6

Accepted: 14 January 2004

This article is available from: <http://www.biomedcentral.com/1471-2164/5/6>

© 2004 Watson et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** The recent development of array based comparative genomic hybridization (CGH) technology provides improved resolution for detection of genomic DNA copy number alterations. In array CGH, generating spotting solution is a multi-step process where bacterial artificial chromosome (BAC) clones are converted to replenishable PCR amplified fragments pools (AFP) for use as spotting solution in a microarray format on glass substrate. With completion of the human and mouse genome sequencing, large BAC clone sets providing complete genome coverage are available for construction of whole genome BAC arrays. Currently, Southern hybridization, fluorescent *in-situ* hybridization (FISH), and BAC end sequencing methods are commonly used to identify the initial BAC clone but not the end product used for spotting arrays. The AFP sequencing technique described in this study is a novel method designed to verify the identity of array spotting solution in a high throughput manner.

**Results:** We show here that Southern hybridization, FISH, and AFP sequencing can be used to verify the identity of final spotting solutions using less than 10% of the AFP product. Single pass AFP sequencing identified over half of the 960 AFPs analyzed. Moreover, using two vector primers approximately 90% of the AFP spotting solutions can be identified.

**Conclusions:** In this feasibility study we demonstrate that current methods for identifying initial BAC clones can be adapted to verify the identity of AFP spotting solutions used in printing arrays. Of these methods, AFP sequencing proves to be the most efficient for large scale identification of spotting solution in a high throughput manner.

### Background

Comparative genomic hybridization (CGH) is a technique used to determine regional DNA copy number changes across an entire genome [1]. This is accomplished by co-hybridizing differentially labeled genomic sample and reference DNA to a metaphase chromosome spread of cultured cells. Analysis of the metaphase chromosomes

will reveal regions of amplification or deletion in the sample DNA [2]. This technique is limited to the resolution at which the amplifications and deletions can be detected of approximately 10–20 Mb [1]. The recent development of array based CGH technology has improved the resolution of genomic profiling [3]. This involves the substitution of the target DNA from metaphase chromosomes to selected

DNA segments spotted onto a microarray, where the distance between target segments determines the resolution. Current methods for creating CGH arrays include spotting whole bacterial artificial chromosomes (BAC) DNA, degenerate oligonucleotide primer (DOP) PCR derivatives of BAC DNA, and amplified fragment pools (AFP) of BAC DNA generated by linker mediated (LM) PCR [4-6]. These procedures aim at producing large quantities of DNA from a library of clones, generating spotting solutions with high DNA concentration.

As the printing of microarrays using whole BAC DNA requires large-scale bacterial culturing and is therefore too labour intensive for projects involving large clone sets. Amplification of BAC DNA by PCR circumvents this limitation. DOP PCR is designed to amplify representative fragments of the BAC DNA with degenerate primers in a single step. LMPCR requires restriction enzyme digestion and linker ligation prior to PCR amplification and is more commonly used (Fig. 1A) as it allows linear amplification. Typically, BAC DNA or its amplified derivative is precipitated and resuspended in spotting solvent prior to array printing.

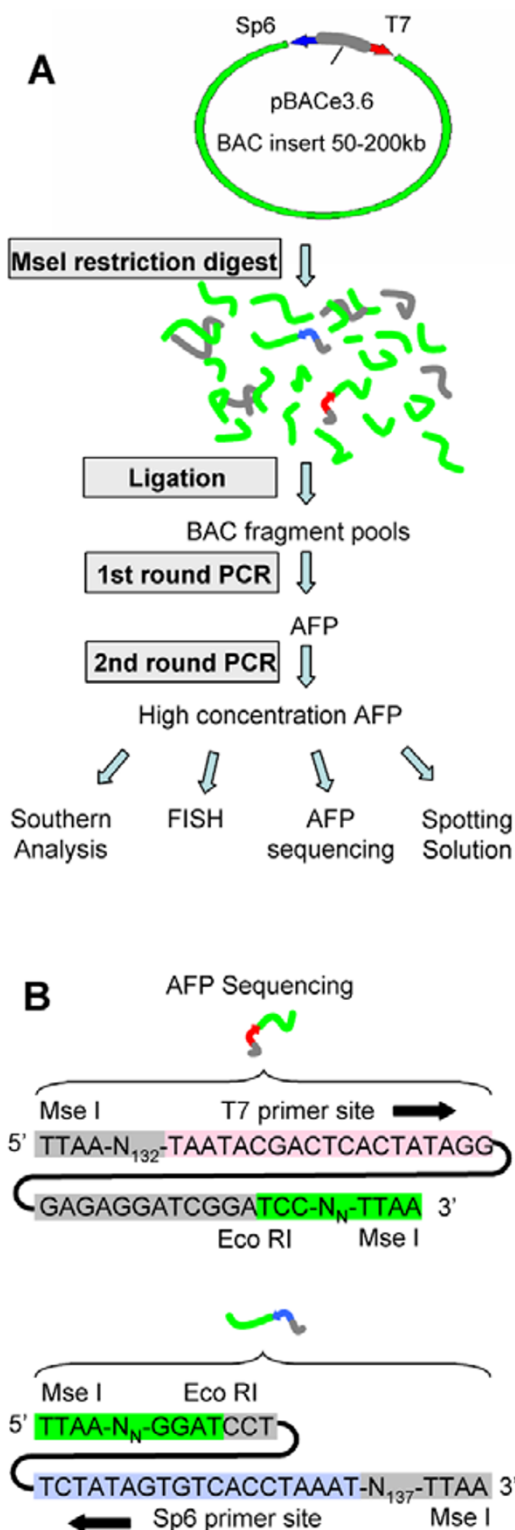
Currently the highest density genome wide CGH array consist of 2460 LMPCR synthesized AFP spaced at 1.4 Mb intervals throughout the human genome [10]. However, with the completion of the human and mouse genome sequencing, large clone sets (tens of thousands of BAC clones) providing complete genome coverage are available for construction of higher resolution arrays [11-14]. Since generating spotting solution from the initial BAC DNA requires multiple liquid transfer steps it is necessary to verify that the final spotting solution is representative of the initial clone. The construction of whole genome arrays necessitates the development of high throughput methods suitable for verification of AFPs prior to spotting arrays.

DNA restriction digest fingerprint analysis, fluorescent in-situ hybridization (FISH) mapping, and BAC end sequencing are commonly used to verify the identity and genomic location of BAC clones [7-9]. However, these clone verification procedures are applied to the BAC DNA prior to multi-step spotting solution synthesis.

Here we demonstrate that these commonly used methods applicable for identification of the initial BAC clone DNA can be adapted for use in verifying AFP just prior to spotting the array.

**Results and discussion**

Southern hybridization, FISH mapping, and modified BAC end sequencing are proven methods for confirming



**Figure 1**  
Flow diagram of analysis of amplified fragment pool (AFP). **(A)** Multistep process for the conversion of BAC DNA to AFP. **(B)** Target fragments for specific primer extension for AFP analysis.

**Table 1: Comparison of three techniques for AFP identification.**

MERITS	FISH ANALYSIS	SOUTHERN ANALYSIS	BAC END SEQUENCING
Chromosomal localization	Genomic position	No	Genomic position
Verification of AFP to original BAC	No	Yes	Yes
Representation of original BAC in AFP	No	Yes	No
Relative cost per assay	Expensive	Moderately expensive	Inexpensive
Time	3+ Days	3 days	4 hours
Other	Requires normal cell line for metaphase	Requires 200 ng of digested BAC DNA	Can be automated for high throughput analysis

clone identity and position in construction of array CGH. In this study we determined if these methods could be applied to verify the amplified fragment pools derived from BAC DNA. The merits of each method are summarized in Table 1.

#### Southern hybridization

Hybridization of the AFP to the *Hind*III digested BAC clone allowed accurate identification (Fig. 2). For example, the AFP derived from the BAC clone RP11-104F13 hybridized to the correct BAC detecting all *Hind*III fragments showing complete representation but did not hybridize to the RP11-104F14, excluding the common vector bands (Fig. 2C). However, in the absence of Cot-1 DNA the AFP cross hybridized to multiple fragments on the wrong clone digest due to the presence of repetitive elements (Fig. 2B). Southern analysis therefore requires the presence of Cot-1 DNA increasing the cost associated with this assay.

#### FISH mapping

An AFP can be labeled as a probe for fluorescent *in situ* hybridization. Metaphase FISH analysis allowed mapping of the AFP to a chromosomal region but did not provide positive identification (Fig. 3). This raises uncertainty when verifying a large clone set since many AFPs will map to the same genomic location within the resolution of FISH on metaphase chromosomes. One concern is if the BAC contains elements which map to multiple areas in the genome, a BAC may hybridize to multiple chromosomal regions even when Cot-1 blocked.

These methods are suitable for sampling AFPs derived from individual BAC clones. Although multiple FISH or Southern analysis can be performed in parallel, these approaches are not easily adapted for high throughput analysis (Table 1).

#### AFP sequencing

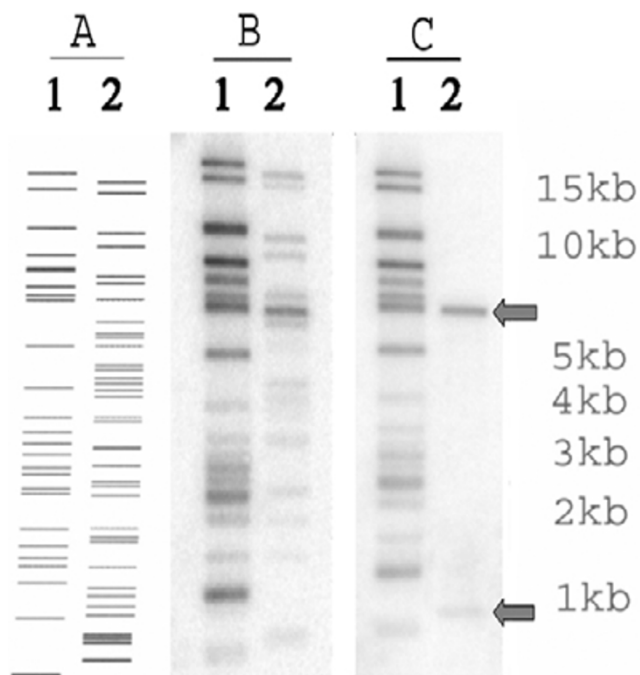
BAC end sequencing can be processed in a 96 well format but requires purified DNA template. AFPs are typically precipitated with ethanol and resuspended directly in

spotting solvent (i.e., 20% DMSO, 50% formamide), which will inhibit the sequencing reaction. In this study we demonstrate that modifications to the Applied Biosystems sequencing protocol allow unpurified AFPs to serve as templates for sequence identification (Fig. 1B). To compensate for sub-optimal conditions due to carry over of unpurified material we increased the template quantity to 20 fmol from the minimum recommended of 2 fmol, and increased the number of sequencing cycles from the typical 35 to 85. Reactions performed using less than 20 fmol or fewer than 85 cycles did not yield sufficient signal for analysis (data not shown). These modifications may have been necessary due to the carry over of primers and reagents from the previous PCR reactions (Fig. 1A) and the complexity of the DNA mixture in the AFP.

To demonstrate the utility of this method we randomly selected 960 clones from the RPCI-11 or RPCI-13 human BAC libraries [15,16] After LMPCR amplification (see methods), 4% of the total unpurified AFP were sequenced using the T7 primer. Half (468) of the AFP yielded sequences and 448 of these were matched to specific BAC clone sequences. Twenty matched repetitive sequences, representing multiple GeneBank entries.

Since the AFPs were generated via a LMPCR protocol involving *Mse*I restriction digested BAC DNA; some of the failed sequence reads may be attributed to the presence of an *Mse*I site downstream of the primer sequence that would truncate primer extension (Fig. 4). To obtain a usable sequence return, the *Mse*I restriction site must be a significant distance from the sequencing primer, preferably greater than 50 nucleotides before *Mse*I recognizes the sequence TTAA.

To determine if the probability of identifying the LMPCR product increased with use of the Sp6 primer, 83 AFPs were sequenced. Of the 83 AFP sequenced, 64 returned usable sequences and 60 of these were matched to a specific BAC. Four matched repetitive sequences, representing multiple GeneBank entries. Combining the results



**Figure 2**  
Use of AFPs as hybridisation probe in Southern Analysis. 200 ng RPII-156K13 *HindIII* digest (lane 1). 200 ng RPII-104F14 *HindIII* digest (lane 2). **(A)** *In silico* fingerprint of RPII-156K13 and RPII-104F14 generated from the BAC database [19] using the FPC software [20]. **(B)** Southern transfer hybridized with radiolabeled AFP from BAC clone RPII-156K13 without Cot-I DNA blocking. **(C)** Southern transfer hybridized with radiolabeled AFP from BAC clone RPII-156K13 with 50 µg Cot-I DNA blocking.

from the Sp6 and T7 sequence reads, it was possible to identify 76 of the 83 AFPs (91%).

## Conclusions

Since PCR amplification of large clone sets are typically processed in a 96 well format, a method for discovering any plate exchanges or mislabeling is essential for quality control of the final AFP set. Of the three methods demonstrated, all identified AFP produced for spotting DNA microarrays. High throughput AFP sequencing will allow identification of 91% of the clones in a clone set when using both the Sp6 and T7 primers. Sequencing of three clones from a plate with the T7 primer allows an 85% determination of plate identity while using Sp6 or both allows 97% and 99.9%, respectively (Fig. 5). For large clone sets the sequencing of all AFPs is desirable but may be prohibitive due to the significant cost associated with large scale sequencing. As a cost effective alternative, we recommend the sequencing of three clones per 96 well

plate for both forward and reverse BAC primers. Direct sequencing of AFPs verified all 96 well plates in our test set. Sequencing of the spotting solution rather than the AFPs is possible only if the spotting solution solvent does not interfere with the sequencing reaction.

The ability to sequence unrefined PCR products and the requirement of only 4% of the AFP makes direct end sequencing of AFP an effective means of verifying array spotting solution.

## Methods

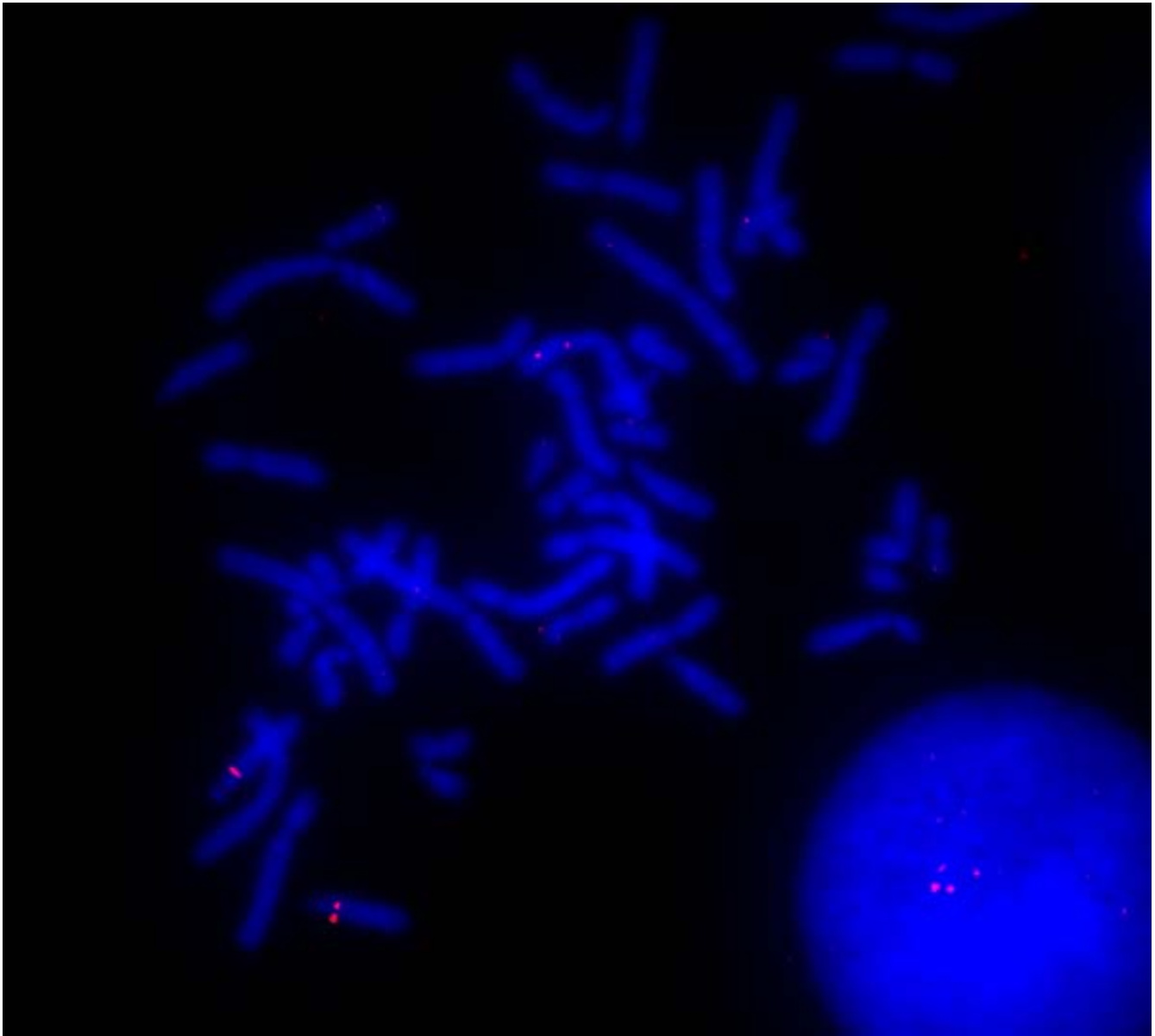
### Linker mediated PCR amplification of BAC DNA

Fifty nanograms of each BAC DNA sample was transferred to a 96 well plate and digested for eight hours with 5 U of *MseI* (New England Biolabs) in a 40 µl reaction. The reaction mixture was inactivated at 65°C for 10 min. Ten percent of the product was transferred to a new plate and ligated to linkers. The ligation mixture consisted of the digested DNA, 0.2 µM primers each of *MseI* long (5' AGT-GGGATTCCGCATGCTAGT 3') and *MseI* short (5' TAACTAGCATCG 3') (Alpha DNA, Quebec) and 80 U of T4 DNA ligase in NEB ligase buffer (New England Biolabs). The primers were allowed to anneal for 5 min at room temperature before addition to the ligation mix. The ligation was performed overnight (12–16 h) at 16°C.

A 2.5 µl aliquot of the 40 µl ligation mixture was amplified in a 50 µl PCR reaction. The reaction mixture contained the linker-ligated DNA template, 8 mM MgCl<sub>2</sub>, 1 mM each dNTP's (Promega), 0.4 µM *MseI* longprimer, and 5 U of Taq polymerase (Promega, storage buffer B) in Promega PCR buffer. After a 3 min 95°C denaturation step, the PCR cycled at 95°C for 1 min, 55°C for 1 min, and 72°C for 3 min, for 30 cycles. A 10 min extension at 72°C completed the protocol. The second round of PCR was initiated using 0.25 µl of the PCR product under the same conditions for 35 cycles. After ethanol precipitation, the final concentration of DNA was quantified using a ND-1000 spectrophotometer (Nanodrop, Delaware). Typical yield for LMPCR was 40–50 µg.

### Sequencing of AFP

To determine the sequence of each amplified fragment pool, 2 µl of AFP was combined with 4 µl Big Dye (Perkin Elmer), 0.32 pmol T7 primer (5' TAATACGACTCACTAT-AGG 3') or SP6 (5' ATTTAGGTGACACTATAG 3') (Alpha DNA) in a 10 µl final volume. After a 1 min initial denaturation step at 95°C, the reaction mixture was subjected to 85 cycles of 95°C 15 s, 50°C for 5 s, and 72°C for 4 min. All steps were ramped at 1°C/s using a MJ Research Peltier thermocycler. The big dye sequencing reaction product was either ethanol precipitated or purified via PCR Min-elute (Qiagen). Sequencing reaction products

**Figure 3**

Use of AFPs as a hybridization probe in FISH analysis. Red represents random primed AFP probe generated from clone RPI I-328P22 (locus: AL353195) labeled with Cy3-dCTP. Chromosomes background stained with DAPI.

were resolved using an ABI Model 377 or ABI Model 3700 sequencer (Applied Biosystems).

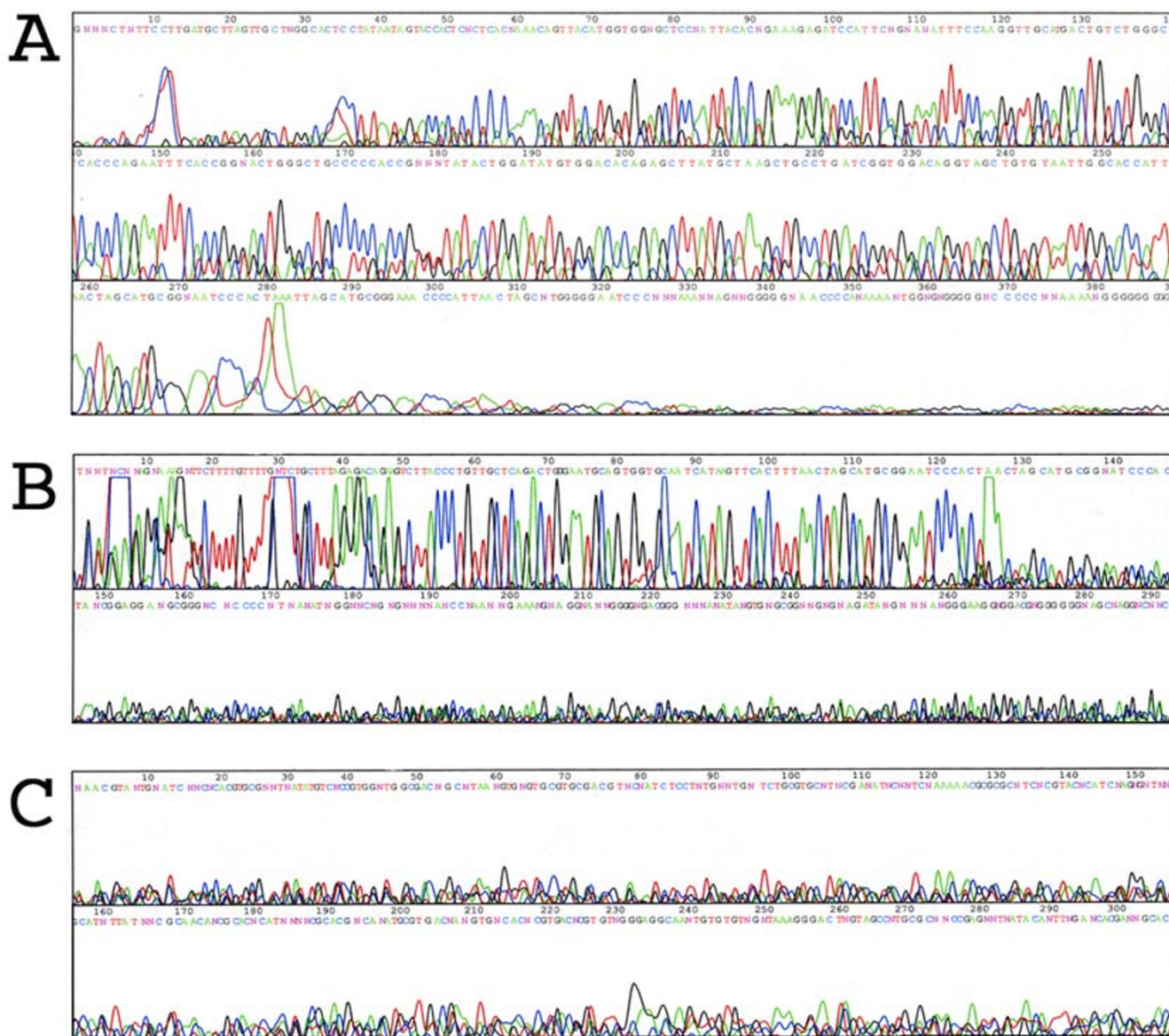
**Clone identification**

Sequences were analyzed using NCBI BLAST to query the non-redundant (nr) and high throughput genomic sequences (htgs) database of GeneBank v.2.2.5. The FTP version of BLAST [17] was downloaded and a script written to allow all 960 sequences to query automatically.

Expect values (E values) of 0.001 and bit scores of 30 were used as the minimum allowed cut off.

**Southern analysis**

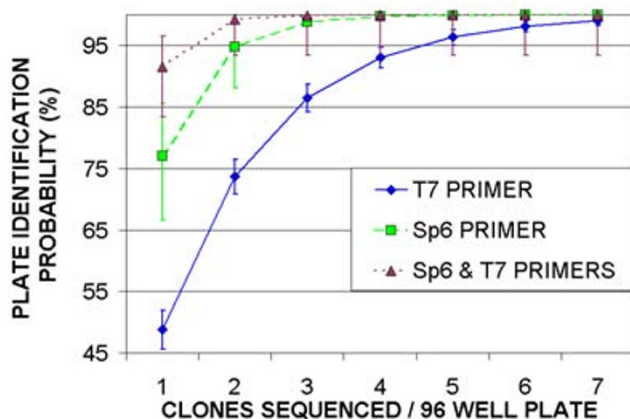
The use of Southern analysis to verify BAC clones for array construction has previously been described [18]. DNA was prepared from overnight cultures of BAC clones. Two hundred nanograms of *Hind*III digested BAC DNA fragments were separated by electrophoresis on a 1% agarose



**Figure 4**  
 Three AFP sequence products. **(A)** Sequence read of an AFP derived from BAC RPI I-124P12 with an *MseI* restriction site 260 bp downstream of the T7 primer. **(B)** Sequence read of an AFP derived from BAC RPI I-125E6 with an *MseI* restriction site 127 bp downstream of the T7 primer. **(C)** Sequence read of an AFP derived from BAC RPI I-124P22 with an *MseI* restriction site 17 bp downstream of the T7 primer.

gel. The separated fragments were transferred to a Hybond-N+ membrane as recommended by the manufacturer (Amersham). One microlitre of AFP (~1 µg) was labeled with α<sup>32</sup>P-dATP using the RadPrime random priming system (Invitrogen). The labeled probes were precipitated in ethanol with (or without) 50 µg Cot-1 DNA (Invitrogen) and redissolved in 15 µl of hybridization

solution (50% formamide, 2X SSC, 10% dextran sulfate, 4% SDS). The probe was denatured at 80°C for 10 min and allowed to cool to 37°C for 2 h before addition to the prehybridized membrane. Hybridization was performed at 65°C overnight in the presence of 0.5 µg/µl of sheared herring sperm DNA (Invitrogen). Washes were performed at 65°C with Buffer 1 (5 mg/ml BSA, 0.5 mM EDTA, 40



**Figure 5**  
Probability of identifying a 96 well plate. In a 96 well format the number of AFP sequenced will increase the probability of identifying the plate. Green solid squares denote Sp6 primer sequencing. Blue solid diamonds denote T7 primer sequencing. Red solid triangles denote sequencing of the AFP with both Sp6 and T7 primers. 95% confidence intervals are represented by vertical bars on all data points.

mM  $\text{Na}_2\text{HPO}_4$  (pH 7.2), 5% SDS) followed by Buffer 2 (2 mM EDTA, 80 mM  $\text{Na}_2\text{HPO}_4$  (pH 7.2), 2% SDS). Autoradiographs were generated from phosphoimager plates and analyzed using the STORM 860 system (Amersham).

#### Fluorescence in situ hybridization

Selected AFPs were mapped by FISH using metaphase chromosomes. Two microlitres of AFP (~2  $\mu\text{g}$ ) were labeled by random priming overnight in the presence of 2 nmol of Cy3-dCTP, Cy5-dCTP (Perkin Elmer), FITC-dUTP, or Texas Red-dUTP using the BioPrime kit (Invitrogen) as per manufacturers directions. The labeled probe was purified using a Sephadex G-50 column, combined with 21  $\mu\text{g}$  of Cot-1 DNA and precipitated with ethanol. The labeled probe was then resuspended in 80  $\mu\text{l}$  of hybridization buffer (50% formamide, 2X SSC, 10% dextran sulfate, 0.1% Tween-20, 10 mM Tris pH 7.4) and denatured for 5 min at 100°C. The metaphase slide was dehydrated through a series of 70%, 80%, and 100% ethanol washes for 2 min each, denatured in 70% formamide in 0.6X SSC for 2 min at 70°C and processed through the same ethanol series at -20°C and allowed to dry. Thirty-five microlitres of probe was then added to the slide and hybridized overnight at 37°C. Images were processed with Qcapture (Q-imaging, Vancouver) with a Zeiss Axio-scope microscope.

#### Authors' contributions

SW performed sequencing analysis and alignment, FISH analysis, and drafted this manuscript. RD participated in the manuscript preparation and southern analysis, AI and CM contributed to sequence analysis and southern analysis. WL is the principle investigator and participated in the design of the study.

#### Acknowledgements

We would like to thank Homa Azad of the BC Cancer Research Centre Sequencing Service and the Michael Smith Genome Sciences Centre at the BC Cancer Agency for performing sequencing reactions. We would like to acknowledge Drs. Donna Albertson and Daniel Pinkel at the University of California at San Francisco for useful discussion on LMPCR methodologies. Also, we would like to thank Bryan Chi for assistance in bioinformatics analysis, Baljit Kamoh for FISH analysis, and Kim Lonergan for manuscript preparation.

#### References

- Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D: **Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors.** *Science* 1992, **258**:818-821.
- Forozan F, Karhu R, Kononen J, Kallioniemi A, Kallioniemi OP: **Genome screening by comparative genomic hybridization.** *Trends Genet* 1997, **13**:405-409.
- Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo W, Chen C, Zhai Y, Dairkee SH, Ljung B, Gray JW, Albertson DG: **High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.** *Nat Genet* 1998, **20**:207-211.
- Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P: **Matrix-based comparative genomic hybridization: Biochips to screen for genomic imbalances.** *Genes Chromosomes Cancer* 1997, **20**:399-407.
- Telenius H, Carter NP, Bebb CE, Nordendkjöld M, Ponder Tunnicliffe BA: **A Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer.** *Genomics* 1992, **13**:718-725.
- Pfeifer GP, Steigerwald SD, Mueller PR, Wold B, Riggs AD: **Genomic sequencing and methylation analysis by ligation mediated PCR.** *Science* 1989, **246**:810-813.
- Cai W-W, Reneker J, Chow C-W, Vaishnav M, Bradley A: **An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization.** *Genomics* 1998, **54**:387-397.
- Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterston RH: **High throughput fingerprint analysis of large-insert clones.** *Genome Res* 1997, **7**:1072-1084.
- Chen X, Knauf JA, Gonsky R, Wang M, Lai EH, Chisoe S, Fagin JA, Korenberg JR: **From amplification to gene in thyroid cancer: a high-resolution mapped bacterial-artificial-chromosome resource for cancer chromosome aberrations guides gene discovery after comparative genome hybridization.** *Am J Hum Genet* 1998, **63**:625-637.
- Snijders AM, Nowak N, Seagraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D, Albertson DG: **Assembly of microarrays for genome-wide measurement of DNA copy number.** *Nat Genet* 2001, **29**:263-264.
- Human Genome Sequencing Consortium: **Initial sequencing and analysis of the human genome.** *Nature* 2002, **420**:520-562.
- Mouse Genome Sequencing Consortium: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520-562.
- Ishkanian A, Watson S, Malloff C, Coe B, DeLeeuw R, Krzywinski M, Marra M, MacAulay C, Lam W: **Construction of a DNA microarray with complete coverage of the human genome [abstract].** *Lung Cancer* 2003, **41**(S2):S60.

14. **Human BAC minimal Tiling Set** [<http://bacpac.chori.org/pHumanMinSet.htm>]
15. Lucito R, Nakimura M, West JA, Han Y, Chin K, Jensen K, McCombie R, Gray JW, Wigler M: **Genetic analysis using genomic representations.** *Proc Natl Acad Sci* 1998, **95**:4487-4492.
16. Frengen E, Weichenhan D, Zhao B, Osoegawa K, van Geel M, de Jong PJ: **A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites.** *Genomics* 1999, **58**:250-253.
17. **The NCBI ftp site** [<http://www.ncbi.nlm.nih.gov/Ftp/index.html>]
18. Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ, de Jong PJ: **A bacterial artificial chromosome library for sequencing the complete human genome.** *Genome Res* 2001, **11**:483-496.
19. **The Human Genome** [<http://www.genome.wustl.edu/projects/human/>]
20. **The Wellcome Trust Sanger Institute** [<http://www.sanger.ac.uk/>]

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

