

# High Resolution Detection Of Chromosome Abnormalities With Single Copy Fluorescence In Situ Hybridization

Peter K. Rogan, Ph.D.<sup>1,2,3</sup> and Joan H.M. Knoll, Ph.D.<sup>1,3</sup>

<sup>1</sup>Childrens Mercy Hospital and Clinics, Schools of Medicine and <sup>2</sup>Computer Science & Engineering University of Missouri-Kansas City, Kansas City MO 64108; <sup>3</sup>Cytogenomix, Overland Park KS, 66213

## ABSTRACT

Precise delineation of rearranged chromosomes in genetic diseases and cancer by fluorescence in situ hybridization can identify the primary etiologies of these disorders. We present a novel approach for defining these abnormalities based on sequential fluorescence in situ hybridization (FISH) of arrays of single copy genomic probes. These arrays consist of sequence-defined synthetic DNA products that can be produced at significantly higher genomic densities than recombinant probes used in conventional FISH. Single copy FISH (scFISH) probes are 50-100 fold shorter than those used for conventional FISH. With scFISH, chromosome abnormalities can be determined at a resolution equivalent to molecular genetic approaches such as Southern analysis. We show how line minimization and bracketing techniques can be used to select probes to optimize strategies for translocation breakpoint determination.

## 1. INTRODUCTION

Precise diagnosis of structural chromosomal abnormalities requires high resolution molecular genetic analyses of patient specimens to identify disrupted genes with altered expression. Breakpoints of intragenic chromosomal rearrangements can be localized to within 20-30 kilobases (kb) using techniques such as Southern hybridization that analyze genomic DNA purified from many cells. In FISH, the location of DNA probe sequences are visualized microscopically on a chromosomal segment or a chromatin fiber in metaphase or interphase [1]. One can identify chromosome segments to correlate chromosome structure with gene location, reveal cryptic abnormalities that are undetectable with conventional banding techniques, and analyze and describe complex rearrangements. In contrast with other molecular genetic techniques, FISH reveals balanced chromosome rearrangements which alter chromosome organization without altering the amount of genetic material as well as mosaic abnormalities that are evident only in a small percentage of cells. FISH had not, until recently, achieved the same degree of genomic resolution as other molecular genetic technologies which quantitatively measure average changes in copy number. Single copy genomic FISH (scFISH) probes can detect chromosome abnormalities at a resolution equivalent to genomic Southern analysis [2]. Chromosomal breakpoints and small chromosomal rearrangements can be visualized and delineated with scFISH probes. We examine the feasibility of developing

these probes genome-wide and discuss probe selection strategies for localizing chromosomal breakpoints at high resolution.

## 2. METHODS

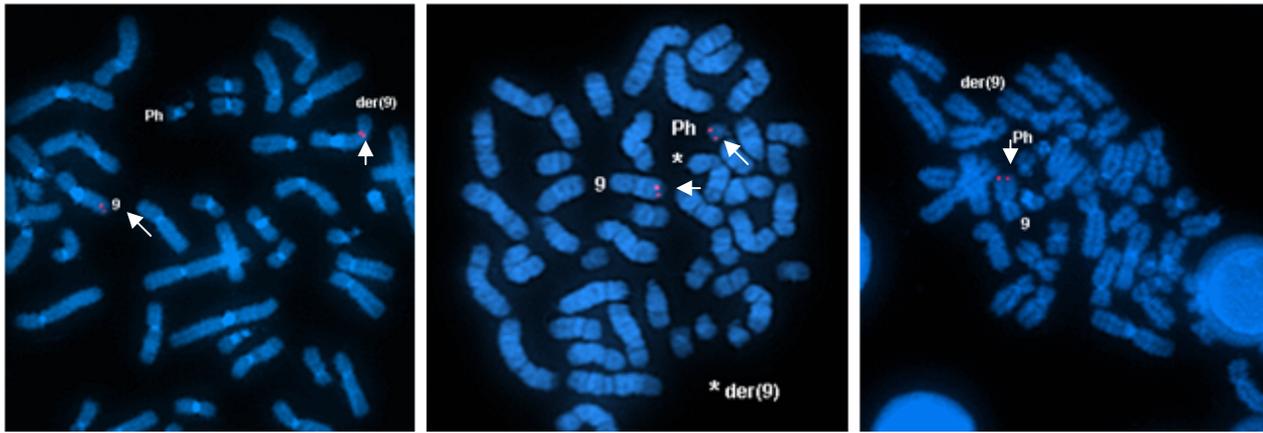
Comparison of a genomic sequence (for example, of a gene that is rearranged in a genetic disease or neoplasia) with a database of previously established sequences of repetitive family members identifies and delineates the coordinates of repetitive elements within the genomic sequence. By excluding the repetitive sequences, probes are designed from >1.5 to 10 kb single copy intervals, synthesized in vitro by long PCR, purified and visualized by fluorescence in situ hybridization to chromosomes (scFISH). Putative probe sequences are then compared with the complete human genome sequence [3] to eliminate those similar to sequences elsewhere in the genome. This approach streamlines the development and production of single-copy, sequence-specific hybridization probes for detection of genetic rearrangements in both rare and common chromosome abnormalities. scFISH probes have been developed and validated from more than 100 chromosomal regions [4, 5].

## 3. RESULTS AND DISCUSSION

### 3.1. Feasibility of genome-wide single copy probe development

To assess the potential for development of scFISH probes throughout the genome, we analyzed the organization of potential single copy probe sequences on chromosomes 21 and 22. Chromosome 21 contains a lower density of genes than chromosome 22, and is somewhat more representative of the complete genome. As expected, chromosome 21 contains fewer single copy intervals than chromosome 22, and the intervals are, on average, shorter. Adjacent intervals tend to be clustered on chromosome 22, with 39% separated by 500-1000 bp. The distributions of interval lengths for both chromosomes are narrow and leptokurtic. Single copy intervals  $\geq 2.3$  kb; the length of a typical scFISH probe are separated, on average, by 29.2 kb on chromosome 21 and by 22.3 kb on chromosome 22.

The centromeric and telomeric regions of chromosome 22 are more densely populated with >2.3 kb intervals than the central region, however the disparity between the different regions is not very marked. The 33 megabase (mb) region close to the telomere has the greatest density (averaging one single copy interval per 8.8 kb), and the 4 mb region close to the centromere has the next highest density (one per 11.1 kb). The



**Figure 1.** Metaphase cells of patients with Chronic Myelogenous Leukemia (CML) and t(9;22) hybridized with short, adjacent single copy FISH probes of 1552 and 1272 bp developed from the genome reference sequence (red spots indicated with arrows). The FISH images show results of hybridization of these probes to three different patients. These sequences are proximal of the ABL1 breakpoint in the top left image, hybridizing to the derivative 9 chromosome. The probes move to the derivative 22 chromosome in the middle panel and are deleted from the derivative 9 chromosome in the right panel.

16-19 mb region in the middle of the chromosome has the fewest single copy intervals (one per 33.3 kb), consistent with its heterochromatic characteristics.

To address the feasibility of developing scFISH probes genome-wide, we first examined the distribution of adjacent single copy intervals on these chromosomes. Most intervals separated by 1.25 kb to 100 kb on chromosome 22q are normally distributed; more densely clustered and sparsely populated chromosomal regions are more prevalent than expected ( $p < 0.0001$ ) and occur throughout the chromosome. We then determined the probability of detecting at least one single copy sequence on chromosomes 21q and 22q. Single copy segments 72 kb in length are found in most 100 kb genomic regions (99% of chromosome 22q; 95% of chromosome 21q). Segments 71 kb are found at least once per 30 kb (>99%). High resolution scFISH should therefore be achievable for most euchromatic regions.

### 3.2 High resolution detection of chromosome rearrangements with scFISH

Molecular cytogenetic techniques have been used to identify the chromosomal source of rearranged, missing, or additional material and to delineate the extent of such chromosomal abnormalities. The primary clinical objective is the identification and delineation of chromosome breakpoints of deletions, insertions or inversions of sequences in a single homolog, and translocations between two or more chromosomes. Conventional FISH, chromosome painting, and array comparative genomic hybridization techniques are suitable for gross identification of abnormal chromosomal material, i.e. chromosome arm, and band assignment, insertions, cryptic translocations, or marker chromosomes. Because of the high density of potential genomic probes, single copy FISH is uniquely suited for applications requiring either precise determination of the boundaries of chromosome rearrangements or the identification of submicroscopic aneuploidies smaller than those detected by conventional recombinant genomic probes. The precision of the scFISH technique can be attributed to the fact that the probes are defined by their coordinates in the human genome reference sequence.

Assays designed to detect chromosome rearrangements that were originally developed for conventional FISH are transferable to scFISH. Either single or multiple distinct probes can be used in a hybridization assay and each probe can be chemically modified so that they can be visualized with different fluorescent labels. We have also developed synthetic probes that hybridize to low-copy, intrachromosomal paralogous sequences in addition to the homologous locus. Simultaneous detection of multiple chromosomal targets with either combinations of single copy probes or through hybridization of a probe to other paralogous sequences can hybridization increase signal intensities and offers additional approaches for detection of chromosomal rearrangements.

Chromosome translocations are among the most common rearrangements in both constitutional (0.1% of the population are carriers) and acquired genetic disorders, such as leukemia. Here, we suggest approaches for localization of chromosome breakpoints in chromosome translocations by scFISH. The genomic context of a probe depends on its proximity and orientation relative to the site of chromosomal breakage. The hybridization experiments can have several different outcomes (Fig. 1). If the probe sequence remains at the original chromosomal location, it is scored as stationary (st). For a sequence that juxtaposed to the translocated chromosome, the probe is defined as having moved (mv) to the derivative chromosome. Hybridization with multiple single copy probes can separate (sep) signals such that at least one probe stays on the original chromosome and at least one moves to the derivative chromosome. A single copy probe that detects one or more additional intrachromosomal paralogous sequences is defined as a split (spl) signal if the breakpoint occurs between the homologous locus and the paralogous copies. In some instances, the chromosomal sequence is missing and the probe is said to be deleted (del).

### 3.3 Optimal strategies for breakpoint refinement of chromosome translocations

The amount of the patient specimen is the primary limitation in localizing chromosomal breakpoints with scFISH, however reagent and labor costs contribute as well. Somewhere between 3

and 20 hybridization experiments can be performed with a single patient sample, depending on the number of cells present and the mitotic index. Thus, the optimal strategy for breakpoint delineation minimizes the number of hybridization assays that are needed.

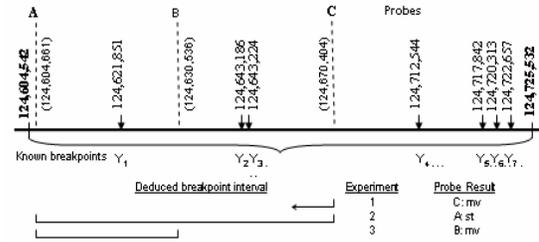
The selection of scFISH probes for chromosome breakpoint detection can utilize nonlinear optimization algorithms for one-dimensional minimization problems (line search) to search a bracketed interval. The boundaries of the interval containing the break may be determined from chromosome positions of the proximal and distal ends of a chromosome band (estimated from the human genome browser; <http://genome.ucsc.edu>), the 5i and 3i terminal coordinates of a gene disrupted in a particular genetic disorder, or from a range of known breakpoints (or breakpoint intervals) that have been defined in previous studies of patients with the disorder. The optimal coordinates of potential probes are then determined using a line search algorithm and the single copy FISH probe(s) that overlaps (or is most proximate to) these chromosome positions is then selected for hybridization.

The distribution of translocation breakpoints in a particular genomic interval (ie. a chromosomal band or a gene) is unknown in patients with *de novo* translocations identified by conventional G banding or in disrupted genes where the sites of rearrangement have not been precisely determined. The number of hybridization assays required for an exhaustive search of an interval defined by coordinates [a, b] at a genomic resolution of  $\epsilon$  base pairs is  $(b-a)/\epsilon + 1$ .

### 3.4 Dichotomous breakpoint search

A bisection bracketing technique was initially used to localize of chromosome rearrangements, which is considerably more efficient than the exhaustive search approach. The hybridization assay determines a Boolean function,  $f(x)$ , whose value is 0 if break is not detected (probe result is st) or is 1 if a break has been detected (result is spl or mv) within the interval [a,b], where sequences ending at position  $i\hat{b}i$  translocate to another chromosome and those at position  $i\hat{a}i$  remain on the original chromosome. The initial coordinates of the probes tested are  $x_1 = a + (b-a)/2 - \epsilon/2$  and  $x_2 = a + (b-a)/2 + \epsilon/2$ . We then compare  $f(x_1)$  and  $f(x_2)$ . If  $f(x_1) < f(x_2)$  then positions  $x > x_2$  are eliminated as potential breakpoints, the interval boundary coordinate  $b$  is reset to  $x_2$ , and new probes with coordinates at (or close to) the recomputed coordinates of  $x_1$  and  $x_2$  are synthesized and hybridized. A similar probe selection procedure is used if  $f(x_1) > f(x_2)$ , except positions of  $x < x_1$  are eliminated and coordinate  $a$  is reset to  $x_1$ . If  $f(x_1) = f(x_2) = 0$ , the breakpoint has not been detected with either probe, so coordinate  $a$  is reset to  $x_2$  and a new initial set of probes is selected by bisection of the revised interval [a,b]. Similarly, if  $f(x_1) = f(x_2) = 1$ , both probes have translocated, and coordinate  $b$  is reset to  $x_1$  and a new initial set of probes is selected within the revised interval [a,b]. The process is continued until the size of the interval,  $b-a$ , is  $< 2\epsilon$ .

We validated this approach by selecting scFISH probes to delineate a breakpoint within intron 1b (IVS1B) of the *ABL1* oncogene on chromosome 9. This gene is disrupted in all patients with chronic myelogenous leukemia and translocated to



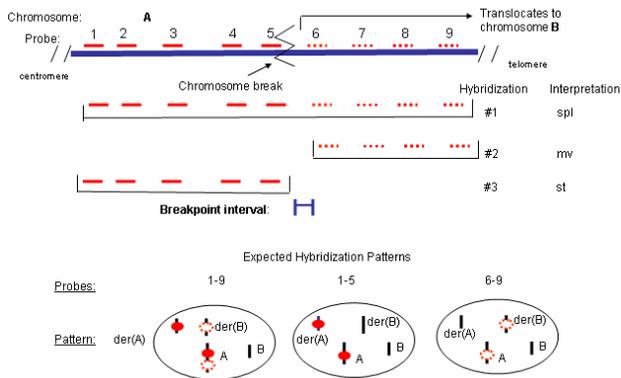
**Figure 2.** Breakpoint localization map in intron 1B of the *ABL1* oncogene in Chronic Myelogenous Leukemia. scFISH probes A, B, and C were used to refine the location of the site of breakage in the coding region of *ABL1* (chromosomal coordinates of April 2003 human genome reference sequence of centromeric end of the probe are given in parentheses; positions of probes are indicated by a dotted line). The results and interpretations of each sequential hybridization experiment are presented below the map. Sites and coordinates of previously determined breakpoints are indicated with arrows. The coordinates of the centromeric and telomeric boundaries of the intron are bolded.

the promoter of the BCR gene on chromosome 22. We have developed 13 scFISH probes from within this 104 kb intron, an average resolution,  $\epsilon$ , of 8 kb. Three of these probes were hybridized sequentially to the same patient specimen (A, B, and C; schematically shown in lower portion of Figure 2). The first hybridization experiment was with probe C which bisected IVS1B. The signal moved (mv) to the derivative 22 chromosome, indicating that the breakage site is upstream of probe C. Since some patients harbor chromosome breaks within exon 1a and IVS1a (which are located upstream of the sequences in Figure 2), hybridization with probe A was performed. The signal remained on the chromosome 9, indicating that the break occurred downstream, within the 66 kb interval delineated by probes A and C. The interval was bisected once again with probe B, which moved to the derivative 22 chromosome, further narrowing the breakpoint to a 26 kb region at the 5i end of IVS1a.

### 3.5 Delineating breakpoints with combinations of probes

Multiple single copy FISH probes that split the hybridization signal onto both the original and derivative chromosomes delineate the upstream (a) and downstream (b) bounds of the translocation breakpoint interval in a single experiment. This type of breakpoint mapping can utilize either interlaced sets of overlapping and unique probes or contiguous, non-overlapping probe arrays in successive hybridization experiments. The initial probes  $x_1 \dots x_n$  are selected based on Golden section partitioning of the initial breakpoint mapping interval [  $x_1 = a + (b-a)*0.382$  through  $x_n = a + (b-a)*0.618$ ]. The algorithm is similar to dichotomous mapping, except in the case of a split signal, which redefines the boundaries of the breakpoint interval based on the terminal coordinates of the most upstream and downstream probes hybridized. Probes are selected by recursive partitioning of the refined interval until the length of the breakpoint interval,  $(b-a)$ , is less than twice the probe resolution.

The translocation breakpoint within the coding region of the *ABL1* gene was localized in another leukemia patient by hybridization of multiple scFISH probes simultaneously (Fig. 3). The initial bracketing interval consisted of the entire 140 kb



**Figure 3.** Strategy for localizing a chromosome translocation breakpoint with a probe array. A contiguous set of scFISH probes are labeled and hybridized first as an ensemble. The signal is split and detects both the original [der(A)] and translocated [der(B)] derivative chromosomes. Subsequent hybridizations with subsets of probes (1-5 and 6-9) that detect der(B) and der(A) exclusive refine the telomeric and centromeric breakpoints, respectively, thus localizing the breakpoint interval. The bottom panel schematically indicates the resulting hybridization patterns seen in scFISH with these probe arrays.

coding region of the gene. Probes developed from IVS3, IVS4-IVS6, and IVS11 all moved to the derivative chromosome 22, which confirmed their location was downstream of the major site of rearrangement within IVS1b. Hybridization with the same 3 probes and 2 additional upstream probes (Probes A and C of IVS1b) revealed split hybridization signals on both the derivative 22 and derivative 9 chromosomes. Therefore, the breakpoint was deduced between IVS1b and IVS3 of *ABLI*. This approach requires fewer hybridization assays to achieve the same degree of breakpoint refinement as the dichotomous approach because only combinatorial probe hybridization can delineate both upstream and downstream boundaries simultaneously. Further improvements in algorithm performance should be achievable by labeling of each probe with different combinations of fluorescent labels. Deconvolution of the emission spectra at the site(s) of chromosomal hybridization should predict which probes define the breakpoint interval and will require fewer hybridization assays.

### 3.6 Weighted probe selection

In many congenital and acquired chromosome abnormalities, the recurrent rearrangements of the same genes have been previously established in other patients with the same diagnosis. In some instances, the sites of chromosomal rearrangement have been precisely determined in one or more individuals, eg.  $y_{i=1,7}$  in Figure 1). This prior information is relevant to the selection of probes that detect recurrent breakpoints in a gene defined by boundaries  $\{a,b\}$ . The cumulative probability distribution of known breakpoints,  $P[x_i > t]$  in which  $\{a > x_i > b\}$ , can guide selection of single copy probes. Chromosome rearrangements are more likely to occur at chromosome coordinates which are local maxima of  $d(P[x_i > t])/dx$ . One potential strategy picks initial probes at the boundaries of breakpoint clusters which are local minima (adjacent to a local maximum) of the first derivative of  $d(P[x_i > t])/dx$ . However, if the breakpoints are not clustered, other Bayesian approaches such as maximum likelihood may be

more appropriate for probe selection, which is suitable for large numbers of established breakpoints. Markov chain Monte Carlo methods may be useful for selecting probes when few breakpoints have been previously determined, a typical situation in many congenital disorders.

There may not be a single optimal solution for this class of probe selection algorithms. In fact, algorithm error may be unacceptable for some abnormalities with sparse, highly dispersed or poorly defined sets of chromosome breakpoints. In such cases, the unweighted dichotomous and combinatorial algorithms may perform better, requiring fewer hybridization experiments for breakpoint localization.

## 4. CONCLUSIONS

scFISH detects and refines the sites of chromosome rearrangement in congenital diseases and neoplasia. scFISH probes can be produced at high densities in regions circumscribing the breakpoints of such rearrangements. Non-linear bracketing techniques can be used to optimize selection of individual or combinations of probes to localize the sequences at chromosomal translocation breakpoints. By accruing sequences of many such breakpoints, it should be feasible to definitively test potential biological mechanisms that underlie these chromosomal rearrangements.

## 5. ACKNOWLEDGEMENTS

We gratefully acknowledge support from the Public Health Service (1R21 CA095167-01), National Science Foundation (Grant # 0228916), Hall Foundation, Paul Patton Charitable Trust, and the Katherine B. Richardson Trust.

## 6. REFERENCES

- [1] Knoll, J.H.M. and Lichter, P, *In situ hybridization to metaphase chromosomes interphase nuclei*, In *Current Protocols in Human Genetics* Vol.1, unit 4.3.(eds.N. C.Dracopoli, et al), John Wiley, New York, 1994.
- [2] P.K. Rogan, P.C. Cazcarro, and J.H.M. Knoll, *Sequence-based design of single-copy genomic DNA probes for Fluorescence In Situ Hybridization*, *Genome Research*, 11, pp. 1086-1094, 2001.
- [3] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, *Basic local alignment search tool*. *Journal of Molecular Biology*, 215, pp. 403-410, 1990.
- [4] J.H.M. Knoll and P.K. Rogan, *Sequence-based in situ detection of Chromosomal Abnormalities at High Resolution*, *American Journal of Medical Genetics*, 121A, pp. 245-57, 2003.
- [5] J.H.M. Knoll, P. Angell, P. Walters, C. Marsh, A. Marion, P. K. Rogan, *Generation of high-specificity, single-copy probes proximate to human telomeres*, submitted.