Am. J. Hum. Genet. 43:934–947, 1988

The Use of Subchromosome-Length Unique Band Sequences in the Analysis of Prophase Chromosomes

David H. Lockwood,*,†,‡ Dennis A. Johnston,† Vincent M. Riccardi,‡ and Stuart O. Zimmerman†

*Graduate School of Biomedical Sciences, University of Texas Health Science Center, †Department of Biomathematics, University of Texas, M.D. Anderson Cancer Center; and ‡Research Cytogenetics Laboratory, Department of Medicine, Baylor College of Medicine, Houston

Summary

Using human prophase chromosome ideograms at the 850-band stage, we previously demonstrated that the 24 prophase ideograms can be divided into a set of 94 unique band sequences, each having a recognizable banding pattern distinct from other nonhomologous chromosome portions. Using actual prophase mitotic cells in this study, we analyzed the p arm of chromosome 11 and of chromosomes 16–22 and characterized a similar set of unique band sequences on actual chromosomes. This set of unique band sequences, a statistical comparison scheme, and image-processing techniques outlined in the present report can be used to identify and distinguish banding patterns of these chromosomes and to determine band pattern abnormalities.

Introduction

Prophase chromosome analyses are utilized in the diagnosis and study of birth defects, in establishing phenotype-karyotype relationships at a refined level, in localizing breakpoints involved in numerous duplication-deficiency rearrangements, and in studying the etiology of neoplasia. Prophase chromosome preparations demonstrate a more complex morphology than metaphase preparations, including extensive chromosome overlapping, bending, and other distortions inherent in their length. This preparation morphology, along with the complex banding patterns, makes prophase chromosome identification and band pattern analysis much more difficult and time consuming than routine metaphase analysis (Schwartz and Palmer 1984). Although seemingly contradicted by this fact, prophase chromosome analysis is often based on routine metaphase chromosome analysis principles involving the identification and interpretation of whole undistorted chromosomes.

Previous experience has suggested that the banding patterns of prophase chromosomes are of sufficient complexity to allow the accurate identification and analysis of subchromosome portions (Lockwood et al. 1983). Chromosome analysis based on subchromosome band pattern recognition and characterization may help minimize the inherent problems of preparation morphology and provide more information on chromosome composition per cell. To test this concept, a search of prophase chromosome G-banding patterns, represented by ideograms (Francke 1981), was undertaken to identify chromosomal segments with banding patterns distinct from those of the rest of the chromosome complement. Ideograms allowed the cytologic problems such as determination of chromosome band stage, disruption of morphology during preparation, and homologue discrepancy to be avoided, focusing solely on band pattern complexity. By means of ideogram profiles and a systematic comparison scheme, the 24 human prophase chromosomes can be divided into a set of 94 unique band sequences, each having a banding pattern recognizable and distinct from any other nonhomologous chromosome portion (Lockwood et al. 1986).

With the establishment of a large number of unique band sequences at the ideogram level, the integrity of

Received June 3, 1987; final revision received August 3, 1988.

Address for correspondence and reprints: Dennis A. Johnston, Department of Biomathematics, Box 237, 1515 Holcombe Boulevard, University of Texas, M. D. Anderson Cancer Center, Houston, TX 77030.

0002-9297/88/4306-0015$02.00

unique band sequences on actual chromosomes was investigated. Since a cytogeneticist's ability to memorize, identify, and utilize efficiently the banding patterns for all 94 unique band sequences is questionable—making a quantitative measure of band pattern similarity or difference desirable—an image-processing approach to chromosome analysis and band pattern comparison was implemented. Image processing approaches to metaphase chromosome analysis began in the 1960s, and current computer-aided metaphase chromosome analysis systems are beginning to reveal their potential as useful tools in chromosome analysis (Neurath et al. 1970; Lundsteen et al. 1986; Philip and Lundsteen 1985). Image processing can digitize images of prophase chromosomes or whole cells and produce density profiles of chromosome or chromosome-segment (unique band sequence) band patterns. The profiles reflect the densities and positions of bands along the chromosome axis (Granlund 1976). Statistical comparison of density profiles gives an indication of profile—and thus of banding pattern, similarity, or difference. This ability to produce and compare profiles allows the cytogeneticist to establish which of the band sequences on actual prophase chromosomes are unique.

The p arm of chromosome 11 and of chromosomes 16–22 were used to test the unique band sequence concept on actual chromosomes. Chromosomes 16–22 were chosen because they are small and easy to digitize and manipulate and because they constitute a good representation of the full spectrum of the banding patterns and chromosomal variation present in the genome. The p arm of chromosome 11 was chosen both because of its previously implied uniqueness (Lockwood et al. 1986) and because of the association of an often subtle interstitial deletion in 11p in patients with aniridia and Wilms tumor (Riccardi et al. 1978). These chromosomes contain 15 of the ideogram-based unique band sequences and account for approximately 18% (by length) of the haploid prophase human genome. The unique band sequences on these chromosomes, as determined at the ideogram level, are shown in figure 1.

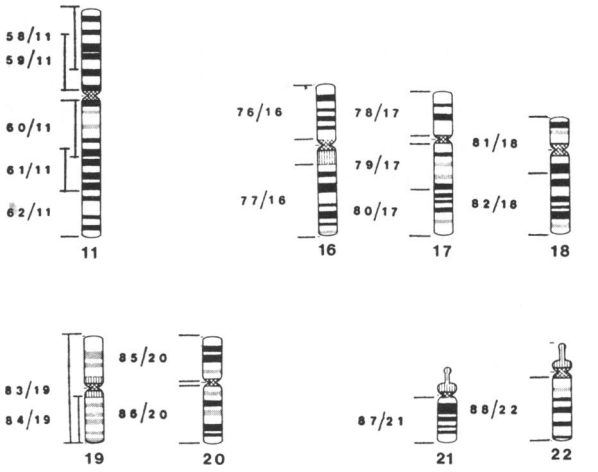

Figure 1 The prophase ideogram unique band sequences for chromosomes 11 and 16 through 22. The numbers at the left of each chromosome represent unique band sequences. These are numbered consecutively beginning with the first sequences located at the p arm of chromosome 1. The second number represents the chromosome number. The extent of each unique band sequence is indicated by horizontal lines. Overlapping vertical lines indicate overlapping unique band sequences. Regions not involved in unique band sequences are not numbered.

Material and Methods

Chromosome Preparations

Prophase chromosome preparations were obtained using a synchronization technique similar to that originally published by Yunis (1976). Phytohemagglutinin-stimulated peripheral blood cultures established in GIBCO 1A culture medium and incubated for 72 h were blocked in S-phase with the addition of methotrexate for 17.5 h. Cells were then released from the block by exposure to a BrdU solution (containing BrdU, FdU, uridine, and deoxycytidine) for 4 h (Riccardi and Holmquist 1979); chromosome condensation was further inhibited by exposure to actinomycin-D during the final hour of culture. Cells were arrested in early mitosis by short exposure (15 min) to colchicine.

Culture harvest, including hypotonic treatment (0.075 MKCl), fixation (3:1 methanol:glacial acetic acid), and slide preparation followed standard procedures. Chromosomes were G-banded using a trypsin-Giemsa method (GTG) (Seabright 1971).

Digitization

Black-and-white 35-mm negatives (Kodak Technical Pan™ film) of prophase chromosome spreads at approximately the 850-band stage were obtained at a magnification of 250× by using a Nikon Optiphot™ HFX microscope and its camera (Nikon Instruments Division, Garden City, NY). Chromosome images on these negatives were scanned using a 1,125-line, high-quality black-and-white Vidicon™ camera (Cohu, San Diego) and were digitized with a Colorado Video 270A digitizer (Colorado Video, Boulder) with joystick

through a parallel interface to a Concurrent/Perkin-Elmer 3230 minicomputer (Concurrent/Perkin-Elmer, Holmdel, NJ) with 2 megabytes of memory. Running on the Concurrent is a command-driven, interactive, general-purpose image-digitizing system performing general image processing and small-object analysis (Johnston et al. 1987). The system consists of 35 first- and second-level overlays containing 47 command/command-systems using a maximum of 335.25 kilobytes of memory. Images and graphics were displayed using a Grinnell Systems GMR 2735 series color-graphic video storage and display system with joystick and a software support package (Grinnell Systems Corporation, San Jose). The Grinnell can display a 512-×-512 three-color pseudocolor or a 6- or 8-bit grayvalue black-and-white image or graphic representation. A Matrix model 2000 color camera (Matrix, Orangeburg, NY) was used for black-and-white and pseudocolor hard copy.

Test chromosomes were digitized at a theoretical resolution range of from 0.27 μm/pixel to 0.04 μm/pixel to determine the optimum resolution or digitization level for prophase chromosomes. This resolution range is at the approximate resolution range of an optical microscope. We determined that an 8-bit gray-level scale, a scan raster of 58.5 pixels/mm of negative image, and 0.07 μm/pixel in the microscopic-image plane provided the spatial resolution necessary for prophase chromosome band recognition. Owing to scan line and digitization overlap, actual resolution was probably closer to 0.2 μm/pixel (approximately the limits of optical resolution). This resulted in a 1920-×- 2048-maximum-pixel cellular image with the representation of an average prophase chromosome 17 being approximately 140 pixels.

Following digitization, a three-step process was required to segment images and allow recognition of single prophase chromosomes or chromosome pieces. First, portions of an image (currently a 512-×-512-pixel maximum size) were displayed on the Grinnell monitor. With a joystick and cursors, touching or overlapping chromosomes of interest were separated by drawing a line segment of low grayvalue (below the detection threshold) between areas to be divided. Next, background level and threshold intensity were determined by black (background)/white (object) plots at possible threshold values. After division and threshold selection, a detection routine separated from the background all objects whose pixel densities were greater than the threshold value and whose cumulative boundary and area sizes were within specified bounds. These objects were then stored separately on a file for future processing. In this manner, individual chromosomes, chromosome pieces, or entire cells could be prepared for further analysis.

Chromosome Profiling

Chromosome profiles are histogram representations of chromosome band patterns, reflecting the densities and positions of bands along the chromosome axis (Granlund 1986). Profiles were produced by integrating the densities along each vector perpendicular to the central axis. Profiles from straight chromosomes or chromosome pieces were easily produced because the axis and perpendicular vectors are clearly defined in the image array. However, the majority of prophase chromosomes displayed a wide variety and complexity of bends. Before profiles could be accurately produced, the central axis and vectors perpendicular to the chromosome axis had to be determined.

An interactive straightening routine was used to straighten all types of bent chromosomes. A chromosome or chromosome piece was displayed on the monitor, and, with a joystick and cursors, centerpoints were defined to estimate the chromosome medial axis. These center points were then used to interpolate the medial axis and chromosome length. With the medial axis and slope of each component line segment thus made known, perpendicular vectors were calculated at each point along the axis, as were interpolated density and position values along each vector through the chromosome outer boundaries. Displayed along a straight line, these vectors resulted in a straightened chromosome image with parallel bands. Although interactive, this algorithm is very efficient and results in accurate chromosome band representations and profiles. The chromosome-straightening routine was utilized whenever dictated by chromosome morphology.

Profile Stretching

Another inherent characteristic of prophase chromosomes is differential homologue length among different prophase cells. Although intercellular and intracellular band patterns at the prophase stage are very similar, overall chromosome length and relative band positions are typically dissimilar. An algorithm to linearly stretch or contract chromosome profiles, a "rubber-band" routine, was applied to digitized profiles to alter the overall chromosome length and relative band positions without affecting the overall pattern of bands. It made in-

dividual chromosomes or whole-cell sets of chromosomes comparable in length, facilitating band-for-band comparisons.

Profile Comparison

The statistical comparison of chromosome profiles was done by a cross-correlation/lagging routine. This routine uses a basic cross-correlation formula described elsewhere (Rosenfield and Kak 1983; Lockwood et al. 1986) to calculate cross-correlation coefficients between any designated chromosome profile and another profile or set of profiles. The designated profile was lagged over the comparison profile in forward and reverse order, pixel by pixel, to achieve all possible pairing combinations. Depending on the maximum correlation value, a determination of profile sameness or difference can be made.

When any two profiles were compared, the "rubberband" algorithm was used in conjunction with the cross-correlation/lagging routine to deal with the presence of homologue contraction/expansion and intercellular chromosome length differences. The profile to which the designated profile was being compared was stretched and contracted, by increments of 3 pixels, from 15% shorter to 15% longer than its original length. These limits of profile expansion and contraction were adequate for typical variations in profile length while minimizing possible profile confusion. The 3-pixel increments were used to minimize computer time. This combination of routine (straightening, stretching, and cross-correlation/lagging) produced the maximum correlation between a designated profile and its comparison profile.

Results

Chromosome Samples

Sample populations of chromosomes at approximately the 850-band stage for each of the chromosomes under study were gathered. These samples did not always comprise whole chromosomes, but, as often was necessitated by the nature of prophase preparations, contained subchromosome segments at least the size of (contained the same number of bands as) the ideogram-based unique band sequences. For each unique band sequence or chromosome being studied, except intact chromosome 19's (sample size = 16) and unique band sequence 81/18 (sample size = 13), the sample size was greater than or equal to 20. Figure 2 shows a representative example of the chromosomes included in the sample. All the chromosomes used were digitized, straightened as necessary, and profiled according to techniques presented in the Material and Methods section.

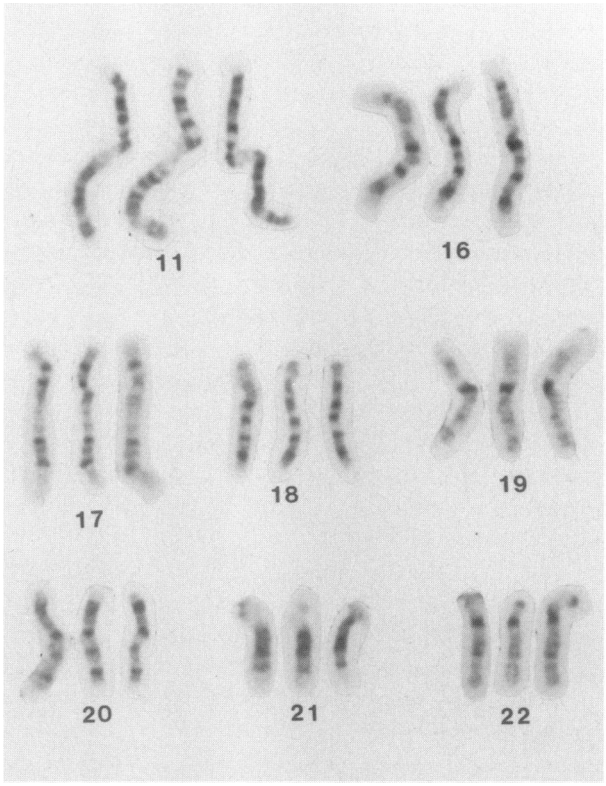

Figure 2 A representative sample of chromosomes 11 and 16 through 23 (three of each) used to establish the presence and application of unique band sequences on actual chromosomes.

Template Unique Band Sequences

For each of the 15 unique band sequences under investigation a template profile had to be developed that mimicked the unique band sequence pattern determined at the ideogram level and had a high correlation with a large percentage of the homologous sample chromosome profiles. First, a correlation value that would serve as the threshold for determining the similarity of or difference between two band pattern profiles was identified. To assess the digitization noise introduced into the chromosome profiles by the imaging system, profiles from 10 digitizations of each of 10 chromosomes were produced. Fischers's Z transformation of these correlation data revealed a lower 95% confidence

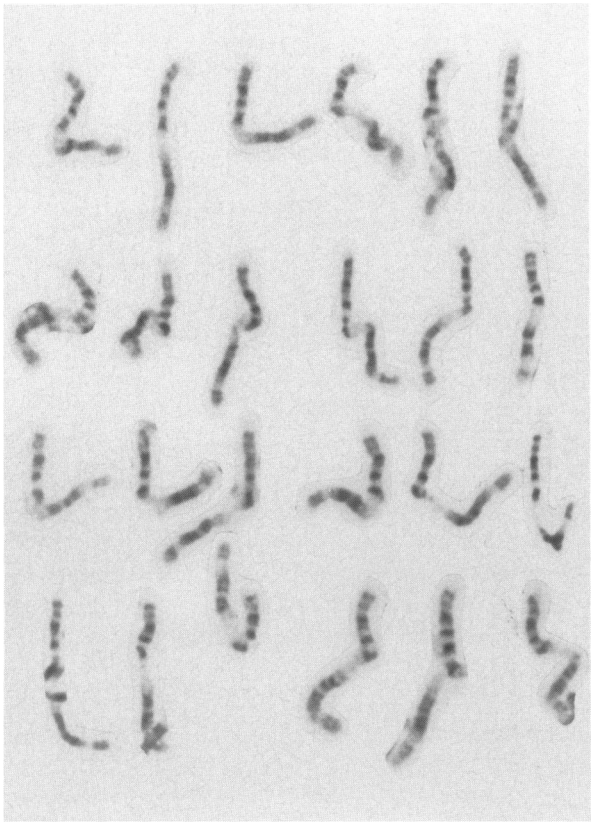

Figure 3 Sample chromosome 11's (approximately 850-band stage)

limit of 0.93 for a hypothesized correlation of .99 (no digitization noise) among these repeat runs (Zar 1984). With a .03 correlation buffer being allowed for chromosomal variation, a correlation of .90 was established as a threshold value to determine profile sameness. Therefore, *any two profiles having a correlation greater than .90 were considered similar, and correlations less than .90 were considered to denote a significant profile difference.*

Figure 3 shows an example of 24 of the 32 chromosome 11's constituting the sample from which the templates for unique band sequences 58/11 and 59/11 were established. These chromosomes were representative of the band pattern similarity at the prophase stage and also of the chromosome variations present, i.e., differences in staining, length, band pattern, and relative position. Figure 4 shows the digitized images of the p arms from these chromosomes, and the resulting profiles are shown in figure 5. Unique band sequence 58/11 included the bands 11p13 through 11p15.5, and unique band sequence 59/11 included the bands 11p11.12 through 11p15.1. In figure 6 is the template for unique band sequence 58/11 (black profile), which had a correlation (defining profile/template sameness) greater than .90 with 26 (81%) of the profile (gray) from the 32 sample 11p's. All 24 chromosomes shown in figure 3 had a correlation greater than .90 with this template. In figure 7 is the template for unique band sequence 59/11 (black profile), which had a correlation greater than .90 with 26 (again 81%) of the profiles (gray) from the 32 samples 11p's. The 26 profiles that matched template unique band sequence 58/11 were not necessarily from the same 26 chromosomes that matched template 59/11. This can be explained both by differential contraction along the chromosome, which may be significant enough to allow some but not all expected unique band sequences to match, and by observing that all 11's were not intact and that some subchromosome segments containing only 58/11 or 59/11 were included.

Similar studies were conducted for the remaining unique band sequences under investigation. Figure 8 shows the templates for unique band sequences 58/11, 59/11, and 76/16 through 88/22. Table 1 shows the

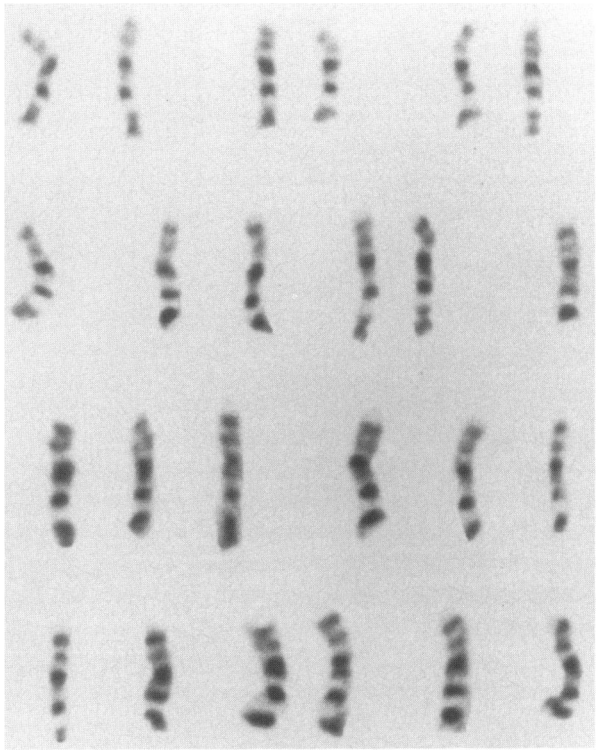

Figure 4 Digitized images from the p arms of chromosome 11's shown in fig. 3.

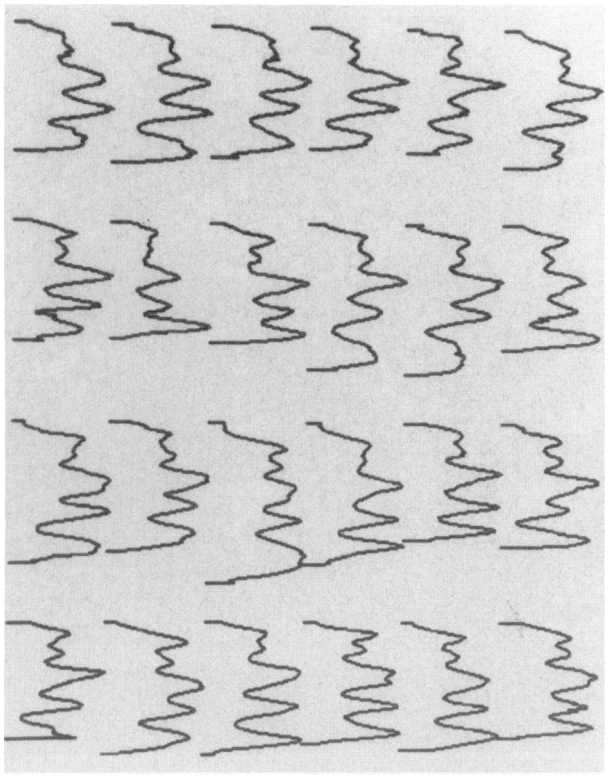

Figure 5 Profiles produced from the digitized images of the p arm of chromosome 11's shown in fig. 4.

sample populations for each chromosome studied and the number and percentage of profiles that their chromosomes matched (correlations greater than .90). No template matched less than 80% of the profiles from its sample chromosome population. The template for unique band sequence 80/17 matched 100% of chromosomes from its sample population, the greatest percentage. Of a total sample of 332 chromosomes or chromosome pieces, 293 (88%) matched their homologous ideogram-determined band pattern templates with a correlation greater than .90. The differences in match percentages may signify chromosome regions of increased variation, poor ideogram representation, poor template construction, or a combination thereof.

Each of the 15 templates representing the ideogram-based unique band sequences under investigation was either an actual chromosome profile from one of the sample chromosomes or an averaging of several profiles that produced the maximum set of correlations to the maximum number of sample profiles. For actual chromosomes, the templates were similar to the banding patterns of the ideogram unique band sequences, except for the templates for unique band sequence 80/17 and 85/20. It was found that on actual chromosomes the unique band sequence 80/17 encompassed bands 17q22 through 17qter. On actual chromosomes, a template mimicking this band pattern had consistently high correlations not only to that banding pattern but also to the banding pattern from 17q21.31 through 17q24.3 when compared in reverse order. A similar situation occurred for unique band sequence 85/20. Therefore, three bands were added to each of these unique band sequences to increase accuracy and improve uniqueness, regardless of orientation.

Template Uniqueness

Having established a set of templates for the unique band sequences on chromosomes 11p and 16 through 22, we had to determine that the templates represented a banding pattern unique to the genome. Complete test-

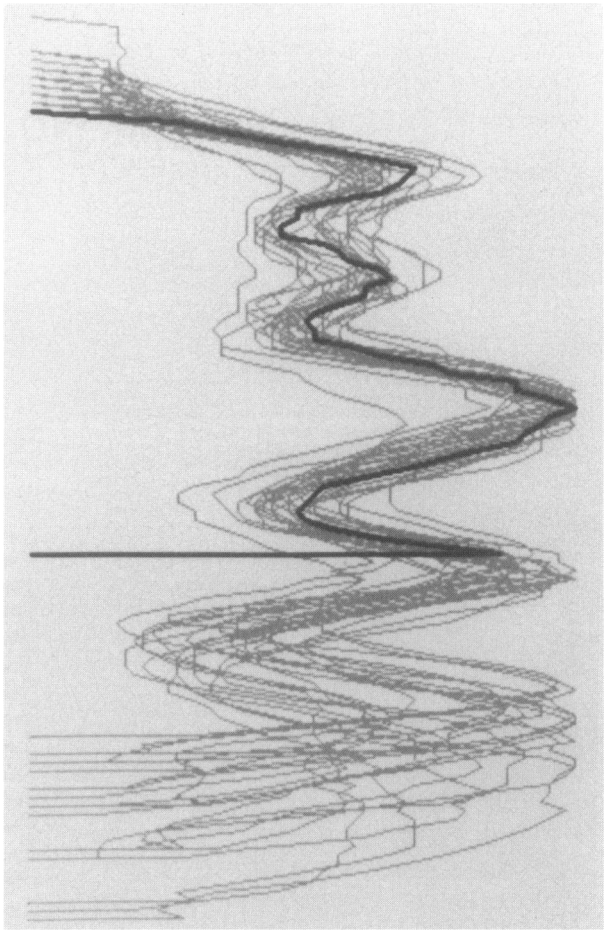

Figure 6 Template for unique band sequence 58/11

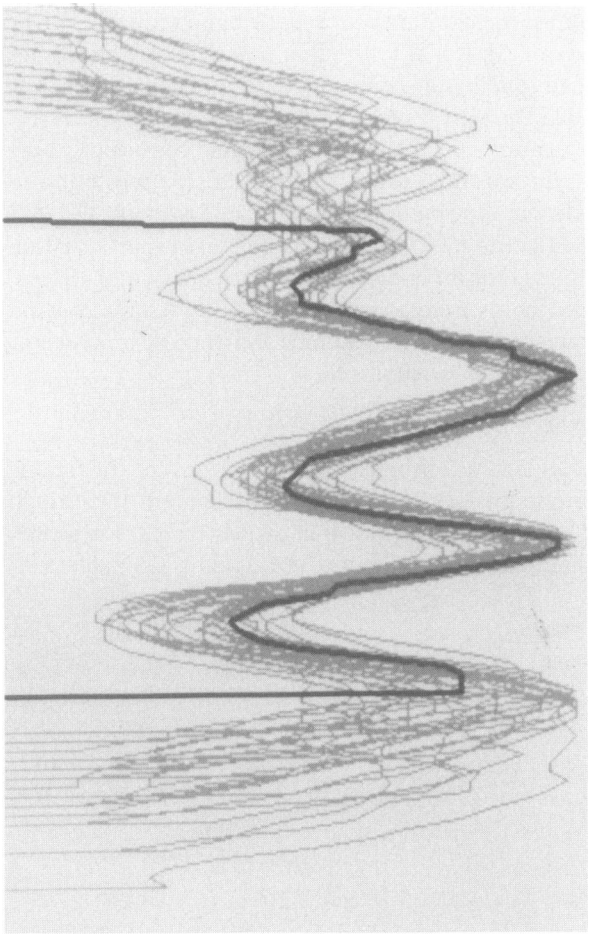

Figure 7 Template for unique band sequence 59/11

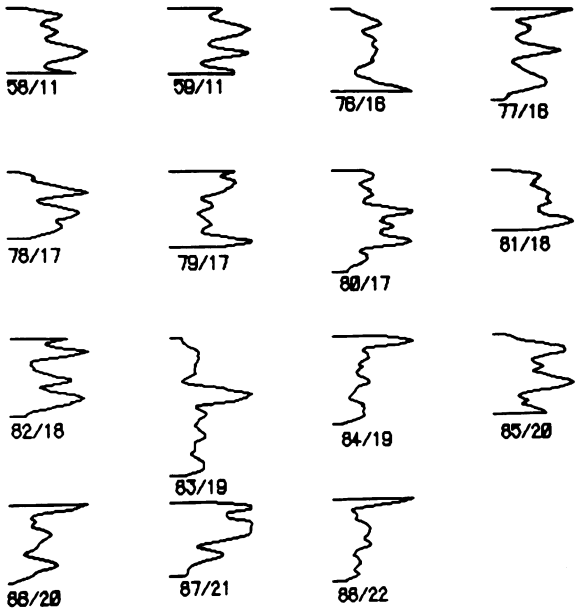

Figure 8 Templates for unique band sequences 58/11, 59/11, and 76/16–88/22 as developed from actual chromosomes.

ing of uniqueness would require adequate sample populations of all the banding patterns in the prophase genome, which was not feasible in this study. However, the potential for genomic uniqueness was demonstrated by comparing each of the 15 templates to the band pattern profiles from all the chromosomes (11p and 16 through 22) in the sample populations. Figure 9 shows

Table 1

Chromosome and Template Comparisons

Chromosome	Sample Size	UBS/CHR Template	Correlations > .90	% Matched
11p	32	58/11	26	81
	32	59/11	26	81
16p	20	76/16	16	80
16q	32	77/16	31	97
17	21	78/17	19	90
	21	79/17	19	90
	20	80/17	20	100
18	13	81/18	11	85
	21	82/18	19	90
19	16	83/19	15	94
	24	84/19	21	88
20	20	85/20	17	85
	20	86/20	19	95
21	20	87/21	17	85
22	20	88/20	17	85
Overall	332	15	293	88

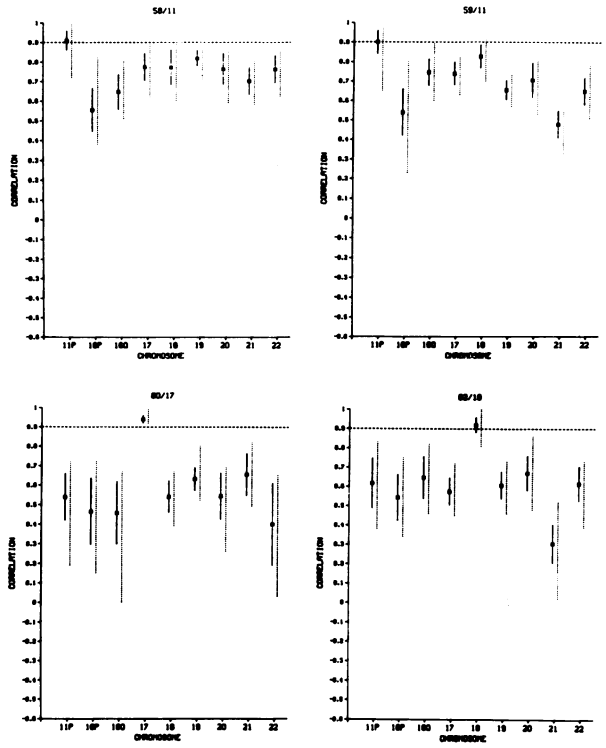

Figure 9 Comparison of unique band sequences 58/11, 59/11, 80/17, and 82/18 with chromosomes 11p and 16 through 22. For each comparison, the solid line and central asterisk represent the correlation mean and SD. Each dashed vertical line represents the range of comparison correlations. The horizontal dashed line represents the .90 correlation cutoff for determining profile and template sameness or difference.

the results of the comparisons between unique band sequences 58/11, 59/11, 80/17, and 82/18 and the sample population of chromosomes. For each template/chromosome comparison, the resulting correlation values are shown on the *y*-axis and the chromosome samples correlated are shown on the *x*-axis. For each comparison group, the solid line and central asterisk represent the correlation mean and SD. Each dashed vertical line represents the range of comparison correlations. The horizontal dashed line represents the .90 correlation cutoff for determining profile and template sameness or difference. For example, the comparison of the template for unique band sequence 58/11 to chromosomes 11p and 16 through 22 revealed that 81% of the profiles from the p arms of chromosome 11 had banding patterns similar to unique band sequence 58/11 (as also was demonstrated in the preceding section). All but one profile (a chromosome 17) from the other chromosomes could be determined to be different from 58/11, demonstrating that template 58/11 was a banding pattern unique to the bands on the p arm of chromosome 11. Similar results are shown for templates 59/11, 80/17, and 82/18.

The data from the testing of all 15 unique band sequences is summarized in table 2, revealing that, with relatively few errors, each template represented a unique banding pattern. Using the correlation value .90 as the threshold to determine profile/template difference, we found that unique band sequence 85/20 had the lowest degree of uniqueness, but it still misclassified only 3% nonhomologous profiles as being similar. The average error rate for the unique band sequences was 1%. Five unique band sequences produced no errors, distinguishing all nonhomologous profiles.

Abnormality Identification

Having shown that a set of unique band sequences similar to those characterized at the ideogram level were also present and identifiable on actual chromosomes, we then applied the concept to the identification of abnormal chromosomes. The identification of a chromosome abnormality is based on the absence of the

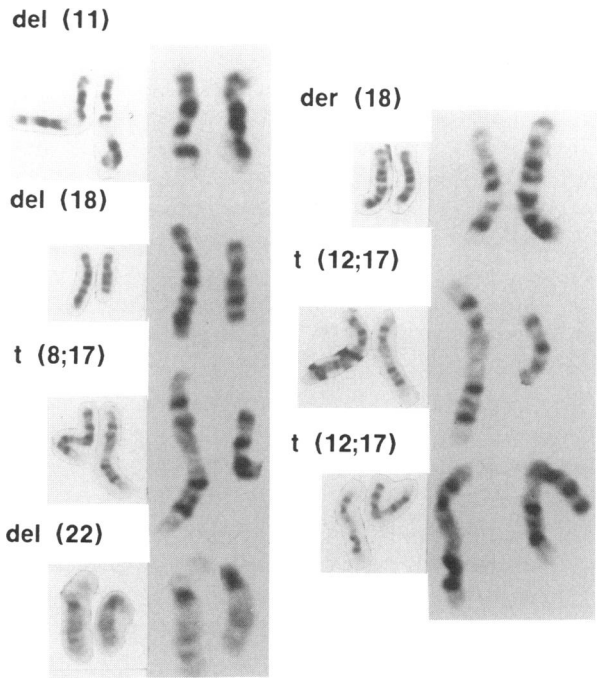

Figure 10 Abnormal chromosomes homologues (*left pair*) and digitized images (*right pair*).

Table 2

Actual Chromosome and Unique Band Sequence Comparisons

Chromosome	Template/Unique Band Sequence						
	58/11	59/11	76/16	77/16	78/17	79/17	80/17
11p:							
Correlation:							
Mean ± SD . . .	.9096 ± .05	.8995 ± .06	.5314 ± .12	.5484 ± .17	.7708 ± .07	.4223 ± .16	.5383 ± .12
Range	.72–1.00	.65–.97	.26–.76	.09–.78	.66–.92	.02–.68	.19–.72
Sample size	(26/32)	(26/32)	(32)	(32)	(1/32)	(32)	(32)
16p:							
Correlation:							
Mean ± SD . . .	.5535 ± .11	.5390 ± .12	.9007 ± .11	.7648 ± .05	.8077 ± .07	.6743 ± .07	.4642 ± .17
Range	.38–.82	.23–.81	.46–1.00	.67–.84	.64–.92	.50–.78	.15–.73
Sample size	(20)	(20)	(16/20)	(19)	(2/20)	(20)	(20)
16q:							
Correlation:							
Mean ± SD . . .	.6452–.09	.7453 ± .07	.7090 ± .07	.9256 ± .04	.7214 ± .07	.5964 ± .12	.4567 ± .16
Range	.51–.81	.60–.90	.58–.84	.74–.98	.57–.86	.37–.86	.00–.68
Sample size	(33)	(1/33)	(33)	(31/32)	(33)	(33)	(33)
17:							
Correlation:							
Mean ± SD . . .	.7729 ± .07	.7387 ± .06	.5907 ± .08	.5356 ± .06	.9106 ± .03	.9150 ± .03	.9400 ± .02
Range	.62–.90	.63–.90	.42–.78	.46–.68	.80 –.95	.84–1.00	.92–1.00
Sample size	(1/30)	(29)	(24)	(25)	(19/21)	(19/21)	(20/20)
18:							
Correlation:							
Mean ± SD . . .	.7716 ± .09	.8264 ± .06	.5854 ± .07	.6011 ± .09	.8184 ± .05	.6296 ± .09	.5392 ± .08
Range	.60–.89	.70–.91	.46–.74	.41–.76	.76–.91	.45–.81	.39–.67
Sample size	(20)	(2/24)	(23)	(24)	(1/21)	(24)	(24)
19:							
Correlation:							
Mean ± SD . . .	.8175 ± .04	.6535 ± .05	.7904 ± .07	.5827 ± .12	.6625 ± .10	.5892 ± .09	.6307 ± .06
Range	.70–.87	.57–.74	.63–.90	.35–.77	.50–.88	.44–.77	.52–.81
Sample size	(24)	(21)	(1/24)	(24)	(23)	(24)	(24)
20:							
Correlation:							
Mean ± SD . . .	.7626 ± .08	.7048 ± .09	.7435–.08	.6949 ± .10	.7769 ± .06	.6699 ± .06	.5426 ± .12
Range	.59–.86	.53–.84	.60–.87	.54–.83	.68–.89	.54–.80	.26–.69
Sample size	(24)	(28)	(23)	(23)	(28)	(23)	(23)
21:							
Correlation:							
Mean ± SD . . .	.7000 ± .07	.4778 ± .07	.1322 ± .17	.2040 ± .15	.7308 ± .06	–.0379 ± .23	.6546 ± .11
Range	.58–.81	.33–.57	–.11–.46	.07–.49	.55–.84	–.51–.34	.49–.82
Sample size	(20)	(20)	(18)	(12)	(20)	(20)	(11)
22:							
Correlation:							
Mean ± SD . . .	.7618 ± .07	.6486 ± .07	.8434 ± .04	.6358 ± .11	.7323 ± .07	.5807 ± .12	.3993 ± .21
Range	.62–.85	.51–.79	.75–.92	.38–.79	.59–.89	.44–.82	.03–.66
Sample Size	(20)	(20)	(1/20)	(20)	(20)	(20)	(20)

Note.—If two numbers are given for "Sample size," the first number represents the correlation matches greater than .90 (indicating correct matches or errors depending on template/chromosome match), and the second number is the sample size.

(continued)

Table 2 (continued)

Template/Unique Band Sequence							
81/18	82/18	83/19	84/19	85/20	86/20	87/21	88/22
.7822 ± .07 .64–.89 (32)	.6161 ± .13 .38–.83 (32)		.4576 ± .16 .14–.89 (32)	.6988 ± .13 .43–.92 (1/32)	.6380 ± .14 .34–.85 (32)	.7671 ± .06 .62–.87 (32)	.5463 ± .1 .32–.7 (32)
.7143 ± .06 .54–.82 (20)	.5426 ± .12 .34–.75 (20)		.8071 ± .07 .63–.90 (1/20)	.7781 ± .14 .30–.93 (3/20)	.8134 ± .07 .67–.92 (2/20)	.6059 ± .13 .21–.79 (20)	.7924 ± .0 .66–.9 (1/20)
.7599 ± .07 .59–.87 (33)	.6466 ± .11 .46–.82 (33)		.7525 ± .09 .58–.89 (33)	.7872 ± .09 .53–.92 (3/33)	.8029 ± .08 .52–.92 (2/23)	.5301 ± .10 .29–.72 (33)	.6496 ± .0 .50–.8 (33)
.8323 ± .06 .69–.92 (1/30)	.5740 ± .07 .45–.72 (24)	.5351 ± .09 .38–.74 (21)	.7032 ± .09 .56–.86 (24)	.7798 ± .06 .64–.86 (23)	.6092 ± .07 .47–.77 (25)	.8417 ± .05 .72–.93 (1/26)	.6440 ± .0 .43–.8 (25)
.9116 ± .07 .70–.99 (11/13)	.9194 ± .04 .81–1.00 (19/21)	.4687 ± .13 .23–.67 (16)	.6528 ± .06 .53–.77 (24)	.5976 ± .10 .39–.77 (22)	.6556 ± .10 .40–.83 (24)	.7280 ± .08 .55–.86 (21)	.5846 ± .1 .38–.8 (23)
.8204 ± .05 .75–.91 (2/24)	.6071 ± .07 .46–.74 (22)	.9327 ± .02 .88–.96 (15/16)	.9179 ± .05 .75–1.00 (21/24)	.7316 ± .05 .59–.81 (22)	.7936 ± .05 .66–.89 (23)	.7097 ± .09 .47–.89 (23)	.8157 ± .0 .70–.9 (2/24)
.7817 ± .05 .69–.87 (29)	.6706 ± .09 .48–.86 (24)	.6044 ± .13 .40–.81 (13)	.7527 ± .10 .60–.94 (2/20)	.9094 ± .05 .75–.98 (17/20)	.9268 ± .04 .80–.98 (19/20)	.6707 ± .08 .53–.83 (24)	.7161 ± .1 .56–.9 (1/24)
.7382 ± .10 .43–.83 (20)	.3054 ± .10 .02–.53 (19)		.2461 ± .13 .05–.48 (13)	.5478 ± .14 .31–.86 (19)	.3124 ± .14 .04–.50 (19)	.9019 ± .06 .70–.97 (17/20)	.2800 ± .1 –.05–.5 (19)
.8033 ± .06 .66–.89 (20)	.6163 ± .09 .39–.74 (20)		.8348 ± .04 .74–.89 (20)	.6270 ± .09 .46–.79 (20)	.8071 ± .04 .72–.87 (20)	.6868 ± .11 .45–.87 (20)	.9233 ± .0 .77–.9 (17/20)

* In each grid the first line of data indicates the means and standard deviation for the comparisons, the second line is the range of correlation values, and the number in parenthesis is the sample size. If two numbers are present in the parenthesis, the first number represents the correlation matches greater than 0.90 (indicating correct matches or errors depending on template/chromosome match), the second number is the sample size.

Table 3

Abnormal Chromosome/Unique Band Sequence Comparisons

Karyotype Signature	Chromosome	UBS	Correlation
46,XX,del(11)(p1305p14.2)	Normal 11	58/11	.9463
		59/11	.9335
	Abnormal 11	58/11	.7687
		59/11	.4170
46,XX,t(8;17)(p23.1;p13.1)	Normal 17	78/17	.9157
	Abnormal 17	78/17	.8132
46,XY,t(12;17)(q24.31;p13.3)[a]	Normal 17	78/17	.9498
	Abnormal 17	78/17	.6435
	Normal 17	78/17	.9239
	Abnormal 17	78/17	.8464
46,XX,del(18)(q21.31→qter)	Normal 18	82/18	.9114
	Abnormal 18	82/18	.6363
46,XY,der(18)(18q+)	Normal 18	81/18	.9048
		82/18	.9097
	Normal 18p	81/18	.9085
	Abnormal 18q	82/18	.7005
46,XY,del(22)(q13.31)	Normal 22	88/22	.9118
	Abnormal 22	88/22	.8692

[a] Two cells analyzed.

identification of the normal banding pattern. A set of seven chromosome homologues in which one chromosome of each homologue pair was abnormal were studied. These abnormalities included homologues from chromosomes 11 (1), 17 (3), 18 (2), and 22 (1). Figure 10 shows the original and digitized chromosomes. In table 3 the karyotype signatures and results of the unique band sequence/abnormal chromosome comparison are summarized. In each case the abnormal banding pattern could be distinguished from the normal banding pattern as revealed by a correlation of greater than .90 for the intact and of less than .90 for the altered band pattern.

For example, the unique band sequences 81/18 and 82/18 were compared against profiles from the two chromosome 18 homologues obtained from a cell with the karyotypic signature 46,XY,der(18)(18q+). A correlation greater than .90 resulted from the 81/18 comparison, because both corresponding banding patterns are normal. In the 82/18 comparison, again a correlation greater than .90 was achieved. However, when unique band sequence 82/18 was compared with the abnormal chromosome homologue, in which the corresponding banding pattern is abnormal, a correlation of less than .90 (.701) was obtained. Although the banding pattern from the one chromosome 18 homologue was abnormal, it only affected the band pattern corresponding to unique band sequence 82/18. As a second, slightly different example, the chromosome 11 homologues from a cell with a karyotypic signature of 46,XX,del(11)(p1305p14.2) are also discussed here. In this abnormality the banding patterns corresponding to unique band sequences 58/11 and 59/11 in the abnormal chromosome 11 were both affected, as shown by correlations of less than .90 for both. Sequences 58/11 and 59/11 both produced a correlation greater than .90 when compared with the normal chromosome 11 homologue. Since some of the homologues discussed in this section overlapped or were distorted, not all of the unique band sequences for a given chromosome could be compared; that is, for the chromosome 17 homologues from the 46,XX,t(8;17)(p23.1;p13.1) cell, only the banding patterns corresponding to unique band sequence 78/17 did not overlap or were not distorted, although unique band sequences 79/17 and 80/17 also corresponded to banding patterns on chromosome 17. These data and those from the ideogram-based abnormality testing support our contention that the unique band sequence templates are specific and are sensitive to change.

Whole-Cell Analysis

The unique band sequence concept was also used to analyze *partially* the banding patterns present in one

prophase cell. (Hardware limitations, which will be addressed in the Discussion, prevented the analysis of a larger number of cells.) This analysis was done to demonstrate the applicability of whole-cell analysis by unique band sequences and image processing. The complete analysis of a prophase chromosome sample would require templates for every unique band sequence; knowledge that, in the entire genome, each template was unique; and analysis of several cells. Judgment on chromosome band pattern integrity within a sample population of cells would be based on the repeated identification of homologous unique band sequences for a specific band pattern in any cell. By conducting paired unique band sequence analysis on a sample of cells, information concerning the genomic chromosome constitution would be provided. Consistent absence of the identification of a specific pair(s) of such sequences would indicate absence of the normal banding pattern; analysis of that region would then be in order.

In the cell investigated, 30 chromosomes or chromosome pieces were isolated and compared with the unique band sequences for chromosomes 11p and 16 through 22. Seventeen of these profiles produced correlation greater than .90 to one of the unique band sequences discussed above. Of these 17, 15 banding patterns were correctly identified. The remaining 13 chromosome segments were from regions not represented by this subset of unique band sequences. The cell is shown in figure 11, and the results of the comparison are summarized in table 4. Again, these data are presented solely to illustrate the concept and feasibility of prophase chromosome analysis based on unique band sequence identification. These data demonstrate the ability to identify and distinguish banding patterns at the whole-cell level. The identification rates could be significantly improved if all the unique band sequences could be used in the analysis. This would allow the development of decision tables. If a profile had a correlation greater than .90 to more than one of the unique band sequences, a judgment concerning the best-fit profile could be made, or the profile could be excluded from analysis. Similar judgments concerning unique band sequence/profile similarity in cases of multiple matches greater than .90 could also be based on analyzing the expected unique band sequence and profile physical distance relationship to telomere location or to adjacent unique band sequences.

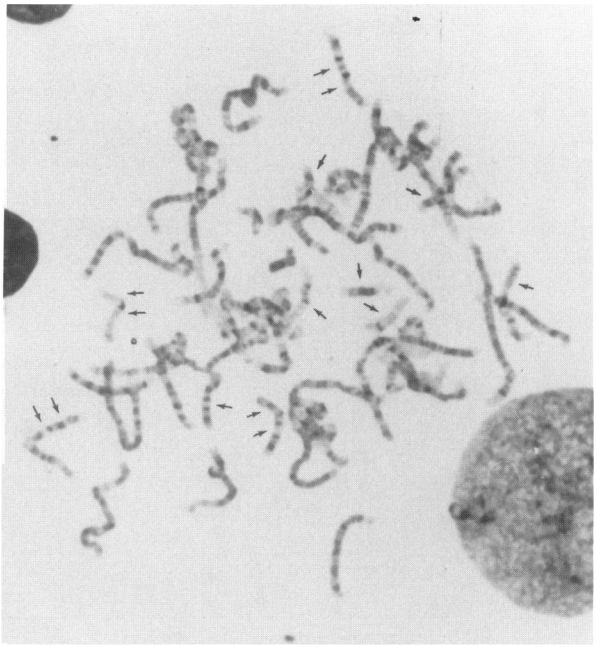

Figure 11 Cell used to demonstrate a whole-cell unique band–sequence approach to prophase chromosome analysis. Arrows indicate correctly identified unique band sequences (with unique band sequences from the p arm of chromosome 11 and chromosomes 16 through 22). The results are summarized in table 4.

Discussion

The major objectives of clinical prophase chromosome analysis are the refinement and localization of abnormalities observed grossly at the metaphase level and the discovery and delineation of abnormalities too minute to be observed in routine metaphase preparations. This high-resolution chromosome analysis requires a precise and meticulous band-by-band characterization of the banding patterns throughout each chromosome. Initial prophase chromosome identification and approaches to such a characterization typically rely on the routine metaphase identification and analysis principles of chromosome length, centromeric index, gross band patterns, and overall chromosome morphology. However, the inherent morphology and complex banding patterns found in prophase chromosome preparations often confound the efficient application of these techniques. With the complex banding patterns and significant levels of artifactual distortion present in prophase chromosome preparations, the necessity and utility of identifying and analyzing whole prophase chromosomes is questionable.

A two-stage approach was taken to identify subchromosome-length unique banding patterns and to apply them to prophase chromosome band pattern identification and characterization. Initially, ideograms were cho-

Table 4

Whole-Cell Analysis

Total Comparisons (N = 30)	UBS	Correlation
Correct comparisons (N = 15)	58/11	.9499
	59/11	.9265
	76/16	.9320
	76/16	.9789
	77/16	.9510
	80/17	.9380
	81/18	.9310
	82/18	.9054
	82/18	.9011
	83/19	.9204
	84/19	.9230
	85/20	.9465
	85/20	.9185
	87/21	.9497
	88/22	.9394
False comparisons (N = 2)	58/11	.9350[a]
	81/18	.9231[b]

[a] Actually a chromosome 10.
[b] Unknown, not a chromsome 18.

sen over actual prophase chromosomes to demonstrate the possible presence of complex and unique band pattern information. In using ideograms, we could ignore cytologic problems such as determining chromosome band stage, disruption of morphology during preparation, and homologue discrepancy and could focus solely on band pattern complexity. Although prophase preparations and ideograms are attainable that have more than 850 bands/haploid human genome, the 850-band level was chosen because techniques for preparing these preparations are routine and readily reproducible. This does not preclude a similar approach to identifying unique band sequences at higher or lower band-stage levels. Previous work demonstrates that, indeed, the ideogram representations of prophase chromosomes at the 850-band stage have band patterns with sufficient inherent information on their own to allow the subdivision of the entire chromosome complement into a set of 94 identifiable unique band sequences.

Following the identification of a set of unique band sequences on the ideograms, this principle was applied to actual chromosomes. The work presented here reveals the presence of a subset of the ideogram-derived unique band sequences (unique band sequences on the p arm of chromosomes 11 and 16 through 22) on actual chromosomes. The combined specificity of the correlation comparison scheme adequately accommodated the adverse effects of preparation morphology and yet allowed subtle variation within and among patterns. The tested subset of chromosomal unique band sequences so closely mimics the ideogram-based unique band sequences that the remaining ideogram unique band sequences for actual chromosomes are probably useful as is. However, the next stage in this research should be the continued refinement of templates for the remaining band sequences and the confirmation on actual chromosomes that each band sequence is unique to the genome.

Following the definition of unique band sequences at the ideogram level and their confirmation against actual chromosomes, they were tested against abnormalities on actual chromosomes. The results show that the unique band sequence concept and comparison scheme are sensitive not only to the corresponding band patterns but also to band pattern alterations. The unique band sequences and comparison scheme distinguished all tested abnormal band patterns from normal band patterns, demonstrating that such a system could identify chromosome deletions and abnormalities involving a single band or less.

To demonstrate the potential for semiautomated prophase chromosome analysis utilizing image-processing techniques and the defined unique band sequences, the ability of unique band sequences 58/11 and 76/16

through 88/22 to identify their respective banding patterns in the partial analysis of a prophase chromosome spread was demonstrated. Our ability to analyze a larger number of cells was limited by available hardware. To analyze whole cells efficiently, digitization and graphics hardware producing a 1,024-×-1,024-pixel resolution (or greater) are necessary. In addition, the image manipulation and comparison algorithms may require array processors and high-speed computers.

With identification of all the unique band sequence templates, complete whole-cell analysis could be conducted. To minimize the error rate of band profile recognition, only recognition of homologue pairs of unique band sequences within a given cell would be accepted. This process would also be conducted not on a single cell but on a population of cells. In addition to correlation data demonstrating unique band sequence/chromosome profile similarity, additional information concerning unique band sequence position, relative to other unique band sequences or telomeres, could be utilized. In cases where two unique band sequences match a given profile, decision tables on similar or commonly confused unique band sequences could be used to help in distinguishing them. In this manner absence of homologue unique band sequence pairs or continued misplacement of pairs would signify abnormal banding regions. This work shows the possibility that the computer can be directed to identify and characterize undistorted unique band sequences in a cell, to repeat this process in a given population of cells, and to derive an accurate report of chromosome constitution based on the sum of findings in all cells. The possibility of computer-assisted prophase chromosome analysis is thus within reach.

At a different level, this information and the procedures discussed here may be useful not only in identifying and characterizing regions of band similarity and possible confusion but also at a more basic level, in exploring chromosome structure and band organization throughout the genome. In addition, this approach and these or similar data may prove helpful in comparative cytogenetics, i.e., quantitatively analyzing and documenting band pattern similarities within and among species.

Finally, as explained elsewhere (Lockwood et al. 1986), this unique band sequence concept can be applied to current methods of prophase chromosome identification and characterization without recourse to an image-processing system. Unique band sequences corresponding to specific areas of concern can be memorized and applied to the analysis of specific band regions (i.e., when analyzing deletion-syndrome studies).

Acknowledgments

This research was supported by grant CA-11430 from the National Cancer Institute, DHHS, and by a Biomedical Research Support Grant from the Graduate School of Biomedical Sciences, The University of Texas Health Science Center at Houston.

References

Francke, U. 1981. High-resolution ideograms of trypsin-Giemsa banded human chromosomes. Cytogenet. Cell Genet. **31**:24–32.

Granlund, G. H. 1976. Identification of human chromosomes by using integrated density profiles. IEEE Trans. Biomed. Eng. BME **23**:182.

Johnston, D. A., M. E. Moffitt, K. S. Tang, D. H. Lockwood, L. E. Newton, and W. B. Cavnar. 1987. An interactive general purpose image digitizing system software manual. 4th ed. Department of Biomathematics, University of Texas M. D. Anderson Cancer Center, Houston.

Lockwood, D. H., V. M. Riccardi, and D. A. Johnston. 1983. Unique band sequences (UBS) of the ISCN prophase chromosome ideogram: a starting place for accurate computer-assisted chromosome analysis (CACA). Am. J. Hum. Genet. **35**:140A.

Lockwood, D. H., V. M. Riccardi, S. O. Zimmerman, and D. A. Johnston. 1986. Prophase chromosome unique band sequences: definition and utilization. Cytogenet. Cell Genet. **42**:141–153.

Lundsteen, C., T. Gerdes, and J. Maahr. 1986. Automatic classification of chromosomes as part of a routine system for clinical analysis. Cytometry **7**:1–7.

Neurath, P. W., M. G. Ampola, D. A. Low, and W. D. Selles. 1970. Combined interactive computer measurement and automatic classification of human chromosomes. Cytogenetics **9**:424–435.

Philip, J., and C. Lundsteen. 1985. Semiautomated chromosome analysis: a clinical test. Clin. Genet. **27**:140–146.

Riccardi, V. M., and G. P. Holmquist. 1979. De novo paracentric inversion in a boy with cleft palate and mental retardation. Hum. Genet. **52**:211–215.

Riccardi, V. M., E. Sujansky, A. C. Smith, and U. Francke. 1978. Chromosomal imbalance in the aniridia-Wilms' tumor association: 11p interstitial deletion. Pediatrics **61**: 604–610.

Rosenfield, A., and A. C. Kak. 1983. Digital picture processing. Academic Press, Orlando.

Schwartz, S., and G. Palmer. 1984. High-resolution chromosome analysis. I. Applications and limitations. Am. J. Med. Genet. **19**:291–299.

Seabright, M. 1971. A rapid banding technique for human chromosomes. Lancet **2**:971.

Yunis, J. J. 1976. High resolution of human chromosomes. Science **191**:1268–1270.

Zar, J. H. 1984. Biostatistical analysis. 2d ed. Prentice-Hall, Englewood Cliffs, NJ.