Final Report
DE-FG06-93ER61553
Chromosome mapping by FISH to metaphase and interphase nuclei
P.I., Barbara Trask
Department of Molecular Biotechnology, University of Washington.
Inclusive dates of support 2/1/93 to 11/31/95; unfunded extension to 9/30/96.

The overall specific aims of this project were:

- 1) to determine the large-scale structure of interphase and metaphase chromosomes, in order to establish new capabilities for genome mapping by fluorescence in situ hybridization (FISH).
- 2) to detect chromosome abnormalities associated with genetic disease and map DNA sequences relative to them in order to facilitate the identification of new genes with disease-causing mutations.
- 3) to establish medium resolution physical maps of selected chromosomal regions using a combined metaphase and interphase mapping strategy and to corroborate physical and genetic maps and integrate these maps with the cytogenetic map.
- 4) to analyze the polymorphism and sequence evolution of subtelomeric regions of human chromosomes.
- 5) to establish a state-of-the art FISH and image processing facility in the department of Molecular Biotechnology, University of Washington, in order to map DNA sequences rapidly and accurately to benefit the Human Genome Project.

We achieved each of these aims during the course of this project. Details of our results are given in the following sections, which are numbered to correspond with the specific aims.

1. Organization of chromatin

This project has contributed to new insight into the organization of human G₀ interphase and metaphase cells. The results defined an appropriate role for FISH in genome analysis. It was well known that DNA is tightly packaged in metaphase cells. We established that DNA sequences can only be reliably ordered along the centromere-telomere axis if they are separated by > 0.5 Mbp, while an even higher threshold may apply at the ends of chromosomes (Trask et al., 1994). Using a combination of interphase (untreated and borate-treated), a correlation between physical distance and genomic separation can be obtained at distances > 50 kbp.

The Trask laboratory has been at the forefront of the study of the structure of interphase chromosome structure in part as a consequence of this DOE funding. Because chromatin is more loosely packed in interphase, resolution of DNA sequences is enhanced to ~ 50 kbp. We made detailed measurements of the distance in interphase between probes separated by 0.05 to 200 Mbp on the same chromosome by taking advantage of mapped probe resources flowing from the human genome project. Our measurements led to a new model of chromosome organization (Sachs et al., 1995; Yokota et al., 1995b). We found that the relationship between mean-square interphase distance and genomic separation has two linear phases with a transition at ~ 2 Mbp. On the first level, DNA appears to be randomly folded as a highly flexible polymer, but is constrained in loops several Mbp in size. This level of organization accounts for the first steep phase of the relationship, a phenomenon that can be exploited to produce detailed genomic maps. We demon-

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

29

MASTER

strated that the parameters of this random-walk, polymer-like structure are different for different regions of the genome, with G-banded regions packed more tightly in interphase than R-banded regions (Yokota et al., 1997b). The second phase of the relationship suggests that the loops are arranged relative to each other along a flexible, randomly folded backbone. Direct evidence was obtained for loops using probes in Xq28, where local decreases in interphase distance with increasing genomic separation was observed, as expected for a loop structure (Yokota et al., 1995b). The simple statistical rules we have established for interphase structure are the basis for the widespread use of interphase FISH mapping to produce high-resolution maps of DNA-sequence probes. Our description of loops places important limitations on the range of distances for which interphase FISH can be used to derive probe order and spacing. The addition of alkaline-borate buffer to nuclei appears to eliminate the higher-level constraints on random-walk chromatin (Yokota et al., 1995a). We found that this treatment significantly extends the range of the first linear phase of the relationship between mean-square interphase distance and genomic separation, providing a way to produce even higher resolution genomic maps.

We have also conducted a large series of experiments that reveal a non-random arrangement of chromosomes within the nuclear volume. These data make it possible to use interphase FISH patterns to detect and characterize chromosome abnormalities, as discussed below. DNA sequence location in the nucleus appears to depend on a) the size of the chromosome on which the DNA sequence is situated, with small chromosomes more centrally located than large chromosomes, and b) the distance of the DNA sequence to its centromere, with centromere-proximal sequences more centrally located than centromere-distal sequences. Two manuscripts describing these findings are currently in preparation (Yokota et al., 1997c; Singer et al., 1997).

2. Detection and characterization of chromosomal abnormalities.

My laboratory has engaged in numerous collaborations with medical geneticists studying chromosome abnormalities associated with specific genetic diseases. These studies were supported at least in part by this DOE grant. In one study, > 45 clones were localized relative to 15 translocation/deletion breakpoints associated with split-hand/split foot disorder on chromosome 7 (Scherer et al., 1994, Ignatius et al., 1996). FISH allowed simultaneous ordering of the YACs and the breakpoints and led to the surprising finding that rearrangements giving rise to the SHSF phenotype are distributed over 500 kbp. The hunt for the gene(s) disrupted in this extensive region is still in progress. This DOE project also supported the following additional collaborations involving similar studies: (a) identification of YACs spanning breakpoints in two incontinentia pigmentosa type I patients and discrimination between these breakpoints using cosmid subclones (Gorski et al., 1996); (c) confirmation of a deletion of PMP22 in hereditary neuropathy with liability to pressure palsy (Rao et al., 1993); (3) demonstration that a recessive form of CMT1A is caused by deletion of one allele and point mutation in the other (ibid.); (f) positioning of YACs relative to 18q- deletions (Silverman et al., 1995; Hui et al., 1995); (g) resolution of an equivocal deletion in a Smith-Magenis syndrome patient by FISH (Juyal et al., 1995); and (h) mapping YACs and BACs relative to overlapping deletions in patients with Williams' syndrome (discussed further below; manuscript in preparation). These diverse studies were only possible because of effective reduction-to-routine practice in the range of FISH techniques.

3. Verification and construction of maps.

FISH continues to play a critical role in the development and verification of genome maps, which in turn form the foundation for expedient large-scale sequencing. We conducted a large

study in collaboration with Eric Green (NHGRI) and Jean Weissenbach (CEPH) to link the physical clone map, the genetic map and the cytogenetic maps of chromosome 7 (Green et al., 1994). Further studies supported in part by this project were directed toward verification and construction of maps and testing of YACs for chimerism by FISH techniques. The genomic regions studied include Huntington region in 4p16.3 (Yokota et al., 1995b); Xq28 (Yokota et al., 1995b, 1997b); the HLA regions of chromosome 6 (Trask et al., 1994; Yokota et al., 1997b); the Alzheimer region on 14q (Bonnycastle et al., 1994); the Werner region on 8p (Yu et al., 1996); and the T-cell receptor beta locus on chromosome 7 (Rowen et al., 1996). FISH was used to map > 175 YACs to chromosomal positions and to test these YACs for chimerism. These studies were done in collaboration with D. Nickerson, P.Y. Kwok, G. Silverman, J. Evans, G. Schellenberg, E. Green, and J. Gorski. Of 33 CEPH mega-YACs analyzed, 45% appeared chimeric. Surprisingly, 15% of 61 YACs derived from a human 7-only hybrid mapped to more than one location on chromosome 7 (Green et al., 1994). It is likely that only some of these YACs are chimeric, while the remainder illustrate that sequences are repeated in different regions of this chromosome.

In addition to these studies, this project allowed us to collaborate with a number of other researchers to map isolated genes of biological significance: the VP16 transcription factor to chromosome X in collaboration with Winship Herr's group (Wilson, et al., 1995); p27kip1 gene to 12p13 in collaboration with Bert Vogelstein's group (Pietenpol et al., 1995); the ACP1 gene to chromosome 2 in collaboration with Robert Van Etten (Bryson et al., 1995); the LAF4 gene to 2q11.2-12 in collaboration with L. Staudt's group (Liao et al., 1996); and the EXTL gene, involved in multiple exostoses with Michael Lovett (Wise et al., 1997).

4. Characterization of duplications in the genome

FISH has identified low-copy repeats as the reason behind anomalous results obtained with other mapping strategies. Our efforts identified (a) a large, very similar duplication involving chromosomes 7 and 9 that complicated Lee Hood's and Lee Rowen's effort to sequence the TCR-beta locus (Rowen et al., 1996); (b) a cluster of repeats within 7q11.2 that are frustrating Eric Green's, Marc Keating's and our efforts to complete this map and identify the genes involved in Williams syndrome; (c) virtually identical duplications spanning ten's of kbp near the telomeres of multiple chromosomes, which hold clues both to the evolution of primates and the diversification of multi-gene families (as well as other duplications detected under other funding support). These results demonstrate that the genome is peppered with large duplications, each of which will frustrate genome-mapping projects that rely on a limited number of methods.

In collaboration with E. Green and M. Keating (University of Utah), we have studied the complex, non-random chromosomal organization of the Williams region in 7q11.23. The aim of this study was to construct a physical map of the region, to correlate patient deletion with phenotypic features, and to identify genes contained within the deleted intervals. Using FISH and over 25 different clones, at least 6 deletion intervals have been defined among 17 patients. The results of STS-content mapping and FISH suggest that the region on either side of the elastin gene consists of a duplicated stretch of DNA present only in this band.

In a separate study, a cosmid known as f7501 has provided an entree into the analysis of a complex subtelomeric region. A large segment of DNA, spanning > 100 kbp) is duplicated, in a polymorphic pattern, at several sites in the genome. The sequence reveals 3 zones of high homology to olfactory receptor genes, raising the possibility that rearrangement of subtelomeres has contributed to the generation of functional diversity of this family. The cosmid's hybridization pattern was studied in 44 individuals from 8 different ethnic populations and non-human primates.

The distribution is remarkable in several respects: a) Three sites, 3q, 15q, and 19q, are labeled in all individuals suggesting that duplications onto these chromosomes occurred long ago; b) at least one individual was found that carries this sequence additionally on other chromosomes (11 chromosomes in total), resulting in polymorphism in location and copy number (7 to 14 copies/cell), c) population differences were observed, suggesting that the distribution of subtelomeric sequences, detectable by FISH, could be used as phylogenetic markers; and d) chimpanzee and gorilla carry only one copy of the region on a chromosome that is not orthologous to any of the human chromosomes, indicating that the sequence has duplicated (or been deleted) during primate evolution. In addition, the high similarity we observe at the sequence level indicates that these subtelomeric repeats may have undergone recent inter-chromosomal gene conversion. A paper describing these findings has been submitted for publication (Trask et al., 1997).

We have also reported on the polymorphic, multi-chromosomal distribution of additional subtelomeric sequences which contribute to the description of these evolutionarily and structurally complex regions of the genome (Martin-Gallardo et al., 1995; Trask et al., 1997).

5. FISH technology

We developed simple FISH techniques to map DNA sequences accurately relative to chromosome bands and each other. Multi-color FISH with cosmid or greater sized probes is routine and therefore can be applied readily to troubleshoot mapping projects. We have integrated different technological approaches for image acquisition and analysis. The need to extract topographic information from thousands of images in a fast and cost-effective manner was accomplished through film-based image collection and the use of a digitizing tablet to enter coordinates of FISH signals. Custom software for this image analysis was written and is available to other researchers. A digital image-processing station was developed, based on IPLab software and custom scripts, to rapidly collect and superimpose -- in registration -- images of multiple fluorochromes. This system greatly enhances our ability to resolve and map relative to chromosomal bands. Several other research groups at UW and Fred Hutchinson Cancer Research Center have made use of our capabilities.

The potential of straightened DNA fibers for mapping was also studied. DNA is straightened by the force of a moving liquid. The DNA sticks to appropriately treated glass once straightened. These DNA molecules can be used to analyze the positions of fluorescently labeled probes or gaps produced by restriction enzymes. A new procedure for reproducibly producing uniformly extended fibers was developed (Yokota et al., 1997a).

A major step toward consolidating advances in FISH technology was the development of a detailed manual describing FISH procedures. This manual was written through support from this project. It is regularly sent out to other laboratories and will soon appear as a chapter in a handbook of genome-analysis methods published by Cold Spring Harbor Press (Trask, 1997).

In summary, we have made significant progress on several fronts, from analysis of nuclear organization to genome mapping to characterization of large-scale polymorphism, as a result of DOE's funding of this project. Cytogenetic tools will continue to play an important role in estimating gaps, orienting and checking contigs, and positioning clones relative to informative chromosome breaks or deletions. Access to this information often vaults a project forward when it has been stalled by the inadequacies of other strategies. Since moving to Seattle, we have worked with ~25 groups who have needed access to the FISH techniques we have developed. Our help has allowed our colleagues to move forward on their mapping, sequencing, and gene-finding endeavors.

ORS.

It is therefore a deep disappointment that DOE has restricted its support of genome-related projects to those projects focused narrowly on DNA sequencing. I am convinced that this strategy is short-sighted. Many of these projects will encounter "problem" areas in the genome -- duplications and polymorphic variation -- that hold clues to evolution, biological function, and human disease, which will require alternative mapping approaches such as FISH. Judging from the flurry of inquiries I have received recently, mapping and sequencing groups in both industry and academia have encountered problems that they hope FISH will solve. It is therefore critical for DOE to continue its support of a few strong groups with proven expertise in this area, such as mine, so that we can respond to the needs of the genomics community.

The following 27 publications were funded in part by this project. They are listed in chronological order:

1. Rao, B.B., Garcia, C.A., Pentao, L., Killian, J.M., Trask, B.J., Suter, U., Snipes, G.J., Ortiz-Lopez, R., Shooter, E.M., Patel, P.I., Lupski, J.R. (1993) Evidence for a recessive PMP22 point mutation in Charcot Marie-Tooth disease type 1A. *Nature Genet.* 5: 189-194.
2. Green, E.D., Idol, J.R., Mohr-Tidwell, R.M., Braden, V.V., Peluso, D.C., Fulton, R.S., Massa, H.F., Magness, C.L., Wilson, A.M., Kimura, J., Weissenbach, J., Trask, B.J. (1994) Integration of physical, genetic, and cytogenetic maps of human chromosome 7: Isolation and analysis of yeast artificial chromosome clones for 117 mapped genetic markers. *Human Molec. Genet.* 3: 489-501.
3. Scherer, S.W., Porrkaj, P., Massa, H., Soder, S., Allen, T., Nunes, M., Geshuri, D., Wong, E., Belloni, E., Little, S., Zhou, L., Becker, D., Kere, J., Ignatius, J., Niikawa, N., Fukushima, Y., Hasegawa, T., Weissenbach, J., Boncinelli, E., Trask, B.J., Tsui, L.-C., Evans, J.P. (1994) Physical mapping of the human split hand/ split foot (SHSF) locus on chromosome 7 reveals a relationship between SHSF and the syndromic ectrodactylies. *Hum. Molec. Genet.* 3: 1345-1354.
4. Trask, B.J., Allen, S., Massa, H., Fertitta, A., Sachs, R., van den Engh, G.J., Wu, M. (1994) Studies of metaphase and interphase chromosomes using fluorescence in situ hybridization. *Cold Spring Harbor Symp. Quant. Biol.* 58: 767-775.
5. Bonnycastle, L.L., Yu, C.-E., Hunt, C.R., Trask, B.J., Clancy, K.P., Weber, J.L., Patterson, D., Schellenberg, G.D. (1994) Cloning, sequencing, and mapping of the human chromosome 14 heat shock protein gene (HSPA2). *Genomics* 23: 85-93.
6. Wilson, A.C., Parrish, J.E., Massa, H.F., Nelson, D.L., Trask, B.J., Herr, W. (1995) The gene encoding HCF, an accessory factor for the herpesvirus transactivator VP16, maps in human Xq28 and is highly expressed in the adult kidney and fetal tissues. *Genomics* 25:462-468.
7. Yokota, H., van den Engh, G., Mostert, M., Trask, B.J. (1995a) Treatment of cells with alkaline borate buffer extends the capability of interphase FISH mapping. *Genomics* 25: 485-491.
8. Silverman, G.A., Schneider, S.S., Massa, H.F., Flint, A., Lalande, M., Leonard, J.C., Overhauser, J., van den Engh, G., Trask, B.J. (1995) The 18q- syndrome: analysis by bivariate

flow karyotyping and the PCR reveals a successive set of deletion breakpoints within 18q21.2-q22.2. *Am. J. Hum. Genet.* 56:926-937.

9. Hui, S.M., Trask, B.J., van den Engh, G.J., Bartuski, A.J., Smith, A., Flint, A., Lalande, M., Silverman, G.A. (1995) Analysis of randomly-amplified flow-sorted chromosomes using the polymerase chain reaction. *Genomics* 26: 364-371.
10. Sachs, R.K., van den Engh, G., Trask, B., Yokota, H., Hearst, J.E. (1995) A random-walk/giant-loop model for interphase chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* 92:2710-2714.
11. Pietenpol, J.A., Bohlander, S.K., Sato, Y., Papadopoulos, N., Liu, B., Friedman, C., Trask, B.J., Roberts, J.M., Kinzler, K.W., Rowley, J.D., Vogelstein, B. (1995) Assignment of the human p27kip1 gene to 12p13 and its analysis in leukemias. *Cancer Research* 55:1206-1210.
12. Yokota H., van den Engh, G., Hearst, J.E., Sachs, R.K., Trask, B.J. (1995b) Evidence for the organization of chromatin in Mbp-sized loops arranged along a random-walk path in the human G0/G1 interphase nucleus. *J. Cell Biology* 130: 1239-1249.
13. Juyal, R.C., Greenberg, F., Mengden, G.A., Lupski, J.R., Trask, B.J., van den Engh, G., Lindsay, E., Christy, H., Chen, K.-S., Baldini, A., Shaffer, L.G., Patel, P.I. (1995) Smith-Magenis syndrome deletion: A case with equivocal cytogenetic findings resolved by fluorescence in situ hybridization. *Am. J. Med. Genet.* 58:286-291.
14. Martin-Gallardo, A., Lamerdin, J., Sopapan, P., Friedman, C., Fertitta, A.L., Garcia, E., Carrano, A., Negorev, D., Macina, R.A., Trask, B.J., Riethman, H.C. (1995) Molecular analysis of a novel subtelomeric repeat with polymorphic chromosomal distribution. *Cytogenet. Cell Genet.* 71:289-295.
15. Bryson, G.L.M., Massa, H., Trask, B.J., Van Etten, R.L. (1995) Gene structure, sequence, and chromosomal localization of the human red cell-type low molecular weight acid phosphotyrosylphosphatase gene, ACP1. *Genomics* 30:133-140.
16. Gorski, J.L., Bialecki, M.D., McDonald, M.T., Massa, H.F., Trask, B.J., Burright, E.N. (1996) Cosmids map two Incontinentia Pigmenti type 1 (IP1) translocation breakpoints to a 180-kb region within a 1.2-Mb YAC contig. *Genomics* 35:338-345.
17. Rowen, L., Koop, B.F., Hood, L. (1996) The complete 685-kilobase DNA sequence of the human beta T cell receptor locus. *Science* 272:1755-1762. (*experiments in Trask lab are acknowledged*)
18. Yu, C.-E., Oshima, J., Hisama, F., Matthews, S., Trask, B.J., Schellenberg, G.D. (1996) A YAC, P1, and cosmid contig and 16 new polymorphic markers for the Werner's syndrome region at 8p12-21. *Genomics* 35:431-440.
19. Liao, X., Ma, C., Trask, B., Massa, H., Gilbert, D.J., Staudt, L.M., Jenkins, N.A., Copeland, N.G. (1996) LAF4 maps to mouse chromosome 1 and human Chromosome 2q11.2-12. *Mammalian Genome* 7: 467-468.
20. Ignatius, J., Knuutila, S., Scherer, S.W., Trask, B.J., Kere, J. (1996) Split hand/split foot malformation, deafness and mental retardation with a complex cytogenetic rearrangement involving 7q21.3. *J. Med. Genet.* 33:507-510.

21. Yokota, H., Johnson, F., Lu, H., Robinson, M., Belu, A.M., Garrison, M.D., Ratner, B.D., Trask, B.J., Miller, D.L. (1997a) A new method for straightening DNA molecules for optical restriction mapping. *Nucl. Acids Res.* 25:1064-1070.
22. Yokota, H., Singer, M.J., van den Engh, G., Trask B.J. (1997b) Regional differences in the compaction of chromatin in human G0/G1 interphase nuclei. *Chromosome Research* 5:1-10.
23. Wise, C.A., Clines, G.A., Massa, H., Trask, B.J., Lovett, M. (1997) Identification and localization of the gene for EXTL, a third member of the multiple exostoses gene family. *Genome Research* 7:10-16.
24. Trask, B.J. (1997) Fluorescence in situ hybridization. In: (Birren, B., Green, E., Hieter, P., Myers, R., eds.) *Genome Analysis: A Laboratory Manual*. Cold Spring Harbor Laboratory Press. (in press, submitted Dec 1994, revised Dec 1995, expected publication of manual: late 1997). (peer-reviewed book chapter).
25. Trask, B.J., Friedman, C., Martin-Gallardo, A., Rowen, L., Akinbami, C., Blankenship, J., Collins, C., Giorgi, D., Iadonato, S., Johnson, F., Massa, H., Morrish, T., Naylor, S., Rouquier, S., Smith T., Wong, D., Youngblom, J., van den Engh, G.J. (1997) Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. (submitted).
26. Yokota, H., Singer, M.J., van den Engh, G.J., Trask, B.J. (1997c) The position of DNA sequences in the human interphase nucleus is dependent on chromosome size and proximity to centromeres (in preparation)
27. Singer, M.J., Yokota, H., van den Engh, G.J., Trask, B.J. (1997) Non-random arrangement of homologous chromosomes in the human interphase nucleus. (in preparation)

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.