LEGIBILITY NOTICE

A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

LA-UR -88-3544

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

LA-UR--88-3544

DE89 002353

TITLE THE NATIONAL LABORATORY GENE LIBRARY PROJECT

AUTHOR(S) Larry L. Deaven and Marvin A. Van Dilla

SUBMITTED TO Hanford Life Sciences Symposium

DISCLAIMER

MASTER

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

The National Laboratory Gene Library Project

Larry L. Deaven
Los Alamos National Laboratory
Life Sciences Division
University of California
P.O. Box 1663, MS M888
Los Alamos, New Mexico 87545

Marvin A. Van Dilla
Lawrence Livermore National Laboratory
Biomedical Sciences Division
University of California
P.O. Box 5507, L-452
Livermore, California 94550

The two National Laboratories at Livermore and Los Alamos have played a prominent role in the development and application of flow cytometry and sorting to chromosome classification and purification. Both laboratories began to receive numerous requests for specific human chromosomal types purified by flow sorting for gene library construction, but these requests were difficult to satisfy due to time and personnel constraints. The Department of Energy, through its Office of Health and Environmental Research, has a long-standing interest in the human genome in general and in the mutagenic and carcinogenic effects of energy-related environmental pollutants in particular. Hence, it was decided in 1983 to use the flow cytometric and molecular biological skills at both laboratories to

construct chromosome-specific gene libraries to be made available to the genetic research community. The National Laboratory Gene Library Project was envisioned as a practical way to deal with requests for sorted chromosomes, and also as a way to promote increased understanding of the human genome and the effects of mutagens and carcinogens on it. The strategy for the project was developed with the help of an advisory committee* as well as suggestions and advice from many other geneticists.

*Membership: P. Berg (Stanford), T. Maniatis and S.A. Latt (Harvard), A.G. Motulsky (U. Washington), W.J. Rutter, (U.C. San Francisco), C.W. Schmid (U.C. Davis), T.B. Shows (Roswell Park), C.T. Caskey (Baylor), F.R. Blattner (U. Wisconsin). M.H. Edgell (U. No. Carolina) and R.E. Gelinas (U. Washington).

The first goal was the construction of small insert (complete digest) libraries for each of the 24 human chromosomal types. These were intended mainly for the medical genetics community for use in the study and diagnosis of genetic diseases by the restriction fragment length polymorphism method. For such studies, short DNA probes on the order of 1 kb in size are ideal, and complete digest libraries are the easiest to produce from the limited amounts of chromosomal DNA available from flow sorting. We made two sets of libraries, one cloned into the Eco R1 site (Los Alamos), and one cloned into the Hind III site (Livermore) of the bacteriophage vector, Charon 21A. A listing of these complete digest libraries is given in Tables 1 and 2. They are available from the American Type Culture Collection, Rockville, MD.

The second goal is the construction of larger insert (partial digest) libraries in phage and cosmid vectors. These libraries will provide a rich source of DNA sequences for physical mapping and for studies of gene structure and function. Our strategy for these constructions is to divide the human karyotype between the two laboratories and to produce one phage and one cosmid library for each chromosome. Our current vector choices include Charon 40 and sCos1 (Los Alamos) and λGEM11 and Lawrst5 (Livermore). These vectors were chosen on the basis of cloning efficiency, sequence preservation, and user convenience. As new and improved vectors become available, they will be considered for use in the project. Partial digest libraries have been made for chromosomes 16, 19, 21, 22, X, and Y. The partial digest libraries will also be available for general distribution through the American Type Culture Collection.

We expect to be able to complete the partial digest libraries in 1.5 to 2 years. Our directions after these libraries are completed are

speculative and to some extent are dependent on changes in technology during this period of time.

At this time, a set of libraries for each mouse chromosome would be a valuable resource to the genetics research community. Our advisory committee had a difficult time in deciding whether the large insert human libraries or mouse libraries should be given first priority for the current work in the project. Depending on changes in technology or activities elsewhere, this need for mouse libraries may or may not exist two years from now. Our expectation is that this will continue to be of high priority, and although the sorting of mouse chromosomes will be difficult, we believe it is feasible. We are currently gaining experience in flow analysis and sorting of mouse chromosomes through collaborative projects with outside investigators. Therefore, construction of mouse chromosome-specific libraries is a likely future direction for the project.

We have also been approached by a group of plant geneticists who would like to see a national effort organized to begin to sort and prepare libraries for a plant species. In this instance, a considerable amount of work will first have to be directed toward developing adequate sources of isolated chromosomes for sorting.

We are also considering the possibility of the construction of sets of human chromosome libraries in one of the new large fragment accepting vectors (YAC, phage P1). Such libraries will probably be necessary for closure of physical maps. Currently, cloning efficiencies would not permit the use of flow sorted chromosomes as a source of target DNA; however, these efficiencies could improve markedly in the coming two years. Other possibilities for the construction of these libraries include the use of monochromosomal hybrids as a source of DNA and the selection of human

clones after library construction. Our ability to monitor hybrids for specific chromosomal content by flow analysis would be advantageous if this approach had to be used.

References

1. Deaven, L.L., Hildebrand, C.E., Fuscoe, J.C., and Van Dilla, M.A., Construction of Human Chromosome Specific DNA Libraries: The National Laboratory Gene Library Project, Genetic Engineering, 8, 317-332, 1986.

2. Van Dilla, M.A., Deaven, L.L. et al., Human Chromosome-Specific DNA Libraries: Construction and Availability. Biotechnology 4, 537-552, 1986.

3. Deaven, L.L., Van Dilla, M.A., Bartholdi, M.F., Carrano, A.V., Cram, L.L., Fuscoe, J.C., Gray, J.W., Hildebrand, C.E., Moyzis, R.K., and Perlman, J., Construction of Human Chromosome-Specific DNA Libraries from Flow Sorted Chromosomes, Cold Spring Harbor Symposium in Quantitative Biology 51, pp 159-168, 1987.

4. Van Dilla, M.A., Deaven, L.L. et al., Chapter III B2 for Flow Cytogenetics, Joe Gray, Ed. (in press).

#38

Table 1

Characteristics of 28 Chromosome-Specific DNA Libraries Cloned into the EcoR1 Site of Charon 21A

Library ID No.[-]	Independent recombinants	Frequency of non-recombinants	Chromosome equivalents*	Chromosome source[+]	Starting Chromosomes ($\times 10^6$)	Starting DNA (ng)
LA 01 NS01	1.3×10^6	<0.01	31	UV24HL10-12	0.5	250
LA 02 NS01	7.0×10^4	0.04	1.8	UV24HL5	0.5	260
LA 03 NS01	2.7×10^4	0.04	0.8	314-1b	0.4	160
LA 03 NS02	2.1×10^5	0.08	6.4	314-1b	0.4	180
LA 04 NS01	5.1×10^4	<0.01	1.6	UV20HL21-27	0.9	370
LA 04 NS02	7.4×10^4	<0.01	2.3	UV20HL21-27	0.9	370
LA 05 NS01	1.3×10^6	<0.01	43	640-12	0.4	130
LA 06 NS01	4.8×10^4	0.06	1.7	UV20HL15-33	1.9	700
LA 07 NS01	2.4×10^5	0.06	9.2	MR3.3:6TG6	0.1	40
LA 08 NS04	3.6×10^4	0.10	1.5	UV20HL21-27	0.5	150
LA 09 NS01	1.6×10^5	0.07	7	HSF-7	0.3	90
LA 10 NS01	4.0×10^5	<0.01	18	762-8A	0.3	90
LA 11 NS02	6.2×10^4	0.17	2.8	80H10	0.4	100
LA 12 NS01	6.0×10^5	<0.01	27	81P5d	1.5	210
LA 13 NS03	7.5×10^4	0.14	4.2	HSF-7	0.4	80
LA 14 NS01	6.1×10^5	<0.01	36	1634	0.5	115
LA 15 NS02	4.0×10^5	0.06	20	81P5d	0.3	71
LA 15 NS03	6.8×10^4	0.09	4	1634	0.4	82
LA 16 NS02	4.0×10^4	0.08	2	HSF-7	3.0	590
LA 16 NS03	3.3×10^5	0.03	21.7	HSF-7	0.5	100
LA 17 NS03	1.1×10^5	0.01	7.9	GM130A	0.6	110
LA 18 NS04	2.5×10^5	0.02	19	HSF-7	0.5	81
LA 19 NS03	1.1×10^5	0.16	11	HSF-7	0.5	65
LA 20 NS01	1.6×10^4	0.24	1.5	HSF-7	0.5	72
LA 21 NS01	1.1×10^6	<0.01	137	HSF-7	0.9	87
LA 22 NS03	9.3×10^4	0.19	11	HSF-7	0.5	55
LA 0X NS01	2.1×10^5	0.02	8.5	81p5d	1.2	380
LA 0Y NS01	1.1×10^5	0.10	11.5	HSF-7	0.5	64

Legends

[-] The ID Code consists of 8 alphanumeric items. The first two items indicate which laboratory made the library, i.e., LL = Lawrence Livermore National Laboratory and LA = Los Alamos National Laboratory. The next 2 items are underlined and indicate chromosome type (in the one case of a mixed 14/15 library, chromosome type is designated as 45). The fifth item is a letter indicating chromosome status, i.e., N for normal, T for translocation, etc. The sixth item is either S (for small insert, complete digest libraries) or L (for large insert, partial digest libraries). The final 2 items represent library construction number.

*Number of recombinants for 1 chromosome equivalent = $\frac{(3 \times 10^9)(0.65)(f)}{4100}$

where: 3×10^9 bp is the size of the human haploid genome; 0.65 is the clonable fraction; f is the fraction of cellular DNA in particular chromosome; 4100 bp is the average fragment size.

[+] Cell lines from which metaphase chromosomes were isolated:

a) Normal diploid human fibroblast lines 761, 811, HSF7

b) Apparently normal human lymphoblastoid line GM130A and multiple X lymphoblastoid line 1634 (49, XXX).

c) CHOx human lymphocyte lines (human chromosome content) —

UV24HL5 (#2, X)
314-1b (#3)
UV20HL21-27 (#4, 8, 21)
640-12 (#5, 9, 12)
UV20HL15-33 (#6, 9, 13, 15, 17, 20, 21)
UV41HL4 (#6, 9)
762-8A (#10)
UV24HL10-12 (#1,3,11,13)

d) E36 X human lymphocyte lines (human chromosome content) 81P5d (X, 12, 15)

e) V79 X human lymphocyte lines (human chromosome content) MR3.3:6TG6 (7 X del

Table 2

Characteristics of 28 Chromosome-Specific DNA Libraries Cloned Into the Hind III Site of Charon 21A

Library ID No.□	Independent recombinants	Frequency of non-recombinants	Chromosome equivalents*	Chromosome source+	Starting Chromosomes (x 10^6)	Starting DNA (ng)
LL 01 NS01	8.3×10^4	0.04	2.1	UV24HL10-12	0.5	270
LL 01 NS02	1.3×10^6	0.01	32 (20)	UV24HL6	0.5	270
LL 02 NS01	6.6×10^5	0.08	16 (5)	UV24HL5	0.5	270
LL 03 NS01	1.6×10^5	0.10	4.8	314-1b	0.5	220
LL 04 NS01	2.3×10^4	0.02	0.8	UV20HL21-27	1.0	400
LL 04 NS02	5×10^5	0.27	10	UV20HL21-27	0.5	210
LL 05 NS01	3.4×10^6	0.26	113 (30)	640-12	0.5	200
LL 06 NS01	7.6×10^5	0.06	27 (20)	UV20HL15-33	0.4	135
LL 07 NS01	3×10^5	0.01	11.5	GM131	0.9	310
LL 08 NS02	2.2×10^6	0.03	93 (20)	UV20HL21-27	0.7	210
LL 09 NS01	3.0×10^5	0.02	13	UV41HL4	0.4	120
LL 10 NS01	2.4×10^5	0.01	10.6	762-8A	0.5	150
LL 11 NS01	1.1×10^5	0.05	4.9	UV20HL4	0.2	50
LL 12 NS01	7.5×10^5	0.01	34 (20)	81P5D	0.4	120
LL 13 NS01	2.2×10^4	0.04	1.3	761	1.0	240
LL 13 NS02	8.5×10^5	0.03	47 (20)	GM131	0.5	120
LL 14 NS01	2.3×10^6	0.06	135	GM131	0.5	110
LL 45 NS01	2.6×10^6	0.02	152 (30)	811	1.0	210
LL 15 NS01	7.0×10^4	0.06	4.4	GM131	1.0	110
LL 16 NS03	7.6×10^5	0.02	51 (20)	HSF7	0.5	100
LL 17 NS02	3.4×10^5	0.02	24 (20)	HSF7	0.5	95
LL 18 NS01	8.9×10^5	0.13	72	761	1.0	170
LL 19 NS01	1.5×10^6	0.02	145 (10)	811	1.0	130
LL 20 NS01	3.9×10^6	0.01	354 (20)	811	1.0	140
LL 21 NS02	4.7×10^5	0.34	60 (20)	811	0.5	45
LL 22 NS01	6.1×10^5	0.05	71 (17)	811	0.5	50
LL 0X NS01	2.1×10^6	0.33	84 (30)	UV24HL5	0.5	170
LL 0Y NS01	2.5×10^5	0.02	27	811	1.0	115

Legends

□ The ID Code consists of 8 alphanumeric items. The first two items indicate which laboratory made the library, i.e., LL = Lawrence Livermore National Laboratory and LA = Los Alamos National Laboratory. The next 2 items are underlined and indicate chromosome type (in the one case of a mixed 14/15 library, chromosome type is designated as 45). The fifth item is a letter indicating chromosome status, i.e., N for normal, T for translocation, etc. The sixth item is either S (for small insert, complete digest libraries) or L (for large insert, partial digest libraries). The final 2 items represent library construction number.

*Number of recombinants for 1 chromosome equivalent = $\frac{(3 \times 10^9)(0.65)/(f)}{4100}$

where: 3×10^9 bp is the size of the human haploid genome; 0.65 is the clonable fraction; f is the fraction of cellular DNA in particular chromosome; 4100 bp is the average fragment size.

Numbers in parentheses refer to representation of amplified library; for very large libraries, only a fraction of the packaging reaction was amplified.

+Cell lines from which metaphase chromosomes were isolated:

a) Normal diploid human fibroblast lines 761, 811, HSF7

b) Apparently normal human lymphoblastoid line GM131

c) CHOx human lymphocyte lines (human chromosome content) —
UV24HL10-12 (#1, 2 del 3, 11, 13, 18)
UV24HL6 (#1, 2del, 3?, 11-13, 14?, 18?, 19)
UV24HL5 (#2, X)
314-1b (#3)
UV20HL21-27 (#4, 8, 21)
640-12 (#5, 9, 12)
UV20HL15-33 (#6, 9, 13, 15, 17, 20, 21)
UV41HL4 (#6, 9, 13, 15, 18, Y)
762-8A (#10, Y)
UV20HL4 (#1, 4-6, 11, 14-16, 19-21)
81P5D (#12, 15, X)