Final progress report

Grant Number: DE-FG02-96ER62258

Title: Fourth International Workshop on Human Chromosome 5

P.I.: John D. McPherson, Ph.D.

This award was administered through the Washington University School of Medicine at St. Louis. The award was for travel funds for U.S. participants to attend the Fourth International Workshop on Human Chromosome 5 that was held in Manchester, November 8-10, 1996.

The following U.S. scientists attended this meeting. A complete list of attendees is in the attached report from the meeting (Appendix A).

Dr. Jan-Fang Cheng,
Human Genome Center,
Lawrence Berkeley Laboratory,
1 Cyclotron Road, MS 74-157,
Berkeley, CA 94720.
United States of America.

Tel: +1-510 486 6590
Fax: +1-510 486 6816
E-mail: jcheng@genome.lbl.gov

Dr. G. Clines,
University of Texas Southwestern, Medical Center,
6000 Harry Hines Blvd.
NB10.106, Dallas,
TX 75235-8591.
United States of America.

Tel: +1-214 648 1676
Fax: +1-214 648 1666
E-mail: clines@mcdermott.swmed.edu

Dr. Michelle Le Beau,
University of Chicago,
Section of Hematology/Oncology,
5841 South Maryland Avenue,
MC 2115,
Chicago, IL 60439.
United States of America.

Tel: +1-773 702 0795
Fax: +1-773 702 0963
E-mail: mmlebeau@mcis.bsd.uchicago.edu

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

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.

Dr. J.D. McPherson,
Department of Genetics,
Genome Sequencing Center,
Washington University School of Medicine,
Box 8501,
4444 Forest Park Blvd.
St. Louis, MO 63108.
United States of America.

Tel: +1-314 286 1848
Fax: +1-314 286 1810
E-mail: jmcphers@watson.wustl.edu

Conover Talbot, Jr.
Genome Database
Johns Hopkins School of Medicine
2024 E. Monument St.
Baltimore, MD 21205-2100.
United States of America.

Tel: +1-410 614 0443
Fax: +1-410 614 0434
E-mail: cct@gdb.org

The main purpose of this meeting was to identify chromosome 5 mapping resources and to integrate these data into a working map and web site. To this end, the meeting focused on individuals with interests in large segments of the chromosome and avoided small regional maps associated solely with a positional cloning project. Also, sequencing of regions of chromosome 5 had been initiated by several centers and the coordination of these efforts was essential.

The meeting participants did successfully prepare an integrated map of the chromosome and delineate regions to be sequenced by each group undertaking this endeavor. Entry of mapping data into GDB was facilitated by the participation of Conover Talbot in the meeting. A Chromosome 5 web site was established as a result of this meeting (http://genome.wustl.edu/gsc/chrom5/Chrom5_home.html). This web site has been maintained to the present. A list of resources for chromosome 5 are maintained at this site with appropriate links to other sites. A copy of the report from this workshop is available at this site as is a copy of the integrated map. The size of the integrated map does not allow it to be readily displayed via a web browser. As such, it is available for downloading at this site.

A copy of the report which was generated from this meeting has been included in Appendix A.

Appendix A

Report of the Fourth International Workshop on Human Chromosome 5 Mapping

Report of the Fourth International Workshop on Human Chromosome 5 Mapping

Prepared by J.D. McPherson and M.J. Dixon

Introduction

The Fourth International Workshop on Human Chromosome 5 was held in Manchester, UK on November 9-10, 1996 and was hosted by the University of Manchester. The workshop was attended by nine participants, whose names, addresses and contact numbers are listed at the end of this report. The workshop was supported by The European Union, through a contract with HUGO, and the Department of the Energy and National Institutes of Health, USA. The major goals of the workshop were: (1) To collate the various genetic, cytogenetic and physical maps of human chromosome 5; (2) To integrate these maps and identify/correct discrepancies between them wherever possible; (3) To catalogue the sequence-ready contigs of the chromosome; (4) To co-ordinate the various sequencing efforts to avoid future duplication; (5) To establish the first (to our knowledge) web site for the human chromosome 5 community which contains the above information in a readily accessible form. Disease genes were not specifically discussed as it was our intention to focus on the generation and integration of chromosome-specific maps.

Brief overviews of the data generated by each laboratory were made by the participants, however, considerable emphasis was placed on the integration of data and maps from various sources, and the electronic submission of these data to GDB.

Genetic maps

A number of laboratories have been involved in the creation of genetic maps incorporating highly short tandem repeat polymorphisms (STRPs). A second generation map was based entirely on STRPs and consisted of 44 markers, 31 of which were positioned with lod 3 support (Weissenbach et al., 1992). This was subsequently updated to provide a map of 137 STRPs, with 64 of these positioned with odds of greater than 1000:1 (Gyapay et al., 1994). Recently, a much higher resolution map of 312 STRPs has been generated with an average interval size of 0.64 cM (Dib et al., 1996). These maps are available electronically (<http://genethon.fr/>). This map was used as the reference map in subsequent map integration efforts of the workshop participants. In addition to the Genethon initiative, maps based on tri and tetranucleotide markers have been produced (Sheffield et al., 1995) and this information is also available electronically (<http://www.chlc.org/>). Other maps that have been published include those generated by EUROGEN (Villamar et al., 1994). Given that the resolution of any genetic map is ultimately limited by the number and distribution of recombination events, the genetic map of chromosome 5 has reached a density where it is becoming inefficient to genotype new markers across the whole range of CEPH families in order to place it on the map. For this reason, a meiotic breakpoint map of chromosome 5 has been compiled. Initially data from various sources was merged to create a genetic map containing 83 STRPs whose positions in the map were supported by odds of at least 1000:1. The average intermarker distance is 2.65 cM. The sex-averaged map length is 219.9 cM (the male map extends over 161.1 cM; the female over 279.7 cM). This map was used to define the meiotic breakpoint panel. The results of this analysis will be available shortly by anonymous ftp (<ftp://gene.ucl.ac.uk/pub/eurogem>).

Physical maps

Since the Third International Chromosome International Workshop on Human Chromosome 5 mapping, the Whitehead Institute has released three versions of their

genome-wide sequence-tagged site content and radiation hybrid maps ((Hudson et al. 1995), <http://www-genome.mit.edu>) These maps formed the framework for the map integration effort of the workshop.

Radiation hybrid maps

A chromosome-specific radiation hybrid map generated at the University of California, Irvine Genome Center was presented by J.D. McPherson. This map contains 556 markers; 134 of these markers are ordered at greater than 1000:1 odds and an additional 114 markers are ordered relative to this first set with odds of between 10:1 and 1000:1. In addition to this framework map, an additional 308 markers have been added to this map and assigned a location based on the two-point data. The framework map includes 75 polymorphic markers for alignment to the genetic map and approximately 60 STSs are in common with those on the Whitehead Institute and Stanford total genome maps. A number of STSs derived from cosmids previously mapped by FISH (Takahashi et al. 1993) were also included to allow alignment of the cytogenetic map.

YAC contigs

A chromosome-wide YAC contig is part of the Whitehead Institute total genome maps. The chromosome 5 YAC map includes 1169 YAC clones assembled into 14 contigs. With increasing emphasis on the use of bacterial-based clones for use in future sequencing efforts a catalogue of the regions which have been cloned in such systems was made. With the exception of those contigs derived from positional cloning efforts there were few sequence-ready contigs available. Large contigs using a combination of cosmids, P1, PAC and BAC clones were presented by several laboratories and these were integrated to provide a resource for sequencing projects.

The Human Genome Center at Berkeley Laboratory has focused its production genomic sequencing on the distal 40 Mb of 5q, because this region contains a cluster of growth factor and receptor genes and is likely to yield new and functionally-related genes through longrange sequence analysis. The mapping strategy used is based on a combination of hybridization and PCR approaches. Inter-Alu fragments generated from non-chimeric YACs covering 3-5 Mb of DNA were used to isolate regionally specific P1/PAC/BAC clones. These clones were sized by pulsed-field gel electrophoresis and their map locations were confirmed using fluorescence in situ hybridization. Approximately 60-70% of the clones representing a targeted region in the P1, PAC or BAC libraries were identified by this approach. Overlaps between P1s, PACs and BACs were mainly established by PCR using STSs generated from ends of clones. Contigs were further orientated using STSs developed from known genes, ordered markers, and ends of P1s, PACs, BACs and YACs. STS content mapping also permitted gap filling by the identification of new clones. Using these combined strategies 198 P1s, 60 PACs and 495 BACs in a 20 Mb region containing the IL4, IL9, FGFA, GRL, ADRB2 and CSF1R genes were mapped. These clones were linked by 563 STSs (mostly derived from clone ends) to form 34 contigs. Of the 34 contigs, 6 are greater than 1 Mb in size, 10 are greater than 0.5 Mb, and 18 are between 0.2 and 0.5 Mb. The same approach has been used to isolate and sub-localize 370 BACs into 74 bins in the distal 20 Mb of 5q3. These clone maps, with STS content information, are distributed through the Lawrence Berkeley Laboratory Web site (<http://www-hgc.lbl.gov/GenomeHome.html>) and will be periodically updated.

As part of their studies into deletions of 5q and malignant myeloidneoplasms, Michele Le Beau's group have defined a commonly deleted segment of 5q31. This region of approximately 4 Mb is flanked by IL9 on the proximal side and D5S166 on the distal side.

A physical map of this interval has been produced. The P1 artificial chromosome (PAC), bacterial artificial chromosome (BAC), and yeast artificial chromosome (YAC) contig consists of 108 clones (78 PAC clones, two BAC, and 28 YAC clones) to which 106 markers (five genes, 11 ESTs, 11 STRPs and 79 STSs) have been ordered. Using PAC clones from the contig for FISH analysis of myeloid leukemia cells with a del(5q), the smallest overlapping deleted segment has been narrowed to approximately 1-1.5 Mb; this region is flanked by PACs P299F9 (proximal) and P98O22 (distal), and spans the interval between D5S479 and D5S500.

Nicholas Fairweather presented data showing that the YAC contig of 5q34-q35 presented by Lu Kuo et al. (1994) had been extended proximally by approximately 15 Mb to include 5q31. Two gaps exist within this region, between D5S119 and SPARC and between FGF1 and D5S207. Seventeen genes, 25 new STSs and 27 anonymous markers were localized to this contig. Furthermore, a cluster of GABAA receptor subunit genes were characterized within this contig. YACs analyzed by FISH and found to be non-chimeric were subcloned into the cosmid vector Supercos1 (Stratagene) to provide six to ten-fold coverage of the genome localized within the YAC contig. The cosmid libraries (35 in total, comprising 12,000 clones) have been replicated into 384 well microtitre dishes and 54 high-density gridded filters prepared. The filters have been used to determine the chimeric status of the cosmid libraries, estimated at less than 4%, and to detect left and right arm clones of the original YACs, and microsatellite- containing cosmids.

Transcript mapping

A gene map of the human genome is maintained at the NCBI (<http://www.ncbi.nlm.nih.gov/SCIENCE96/>) which is a collection of the EST mapping data from an international consortium (Schuler et al. 1996). This searchable catalog currently lists 974 cDNA which have been regionally localized to chromosome 5.

Greg Clines presented a transcript map of chromosome 5 which has been developed as part of the attempts to build a more saturated transcript map using chromosome-specific cDNA (CS cDNA) libraries derived from fetal brain, placenta, thymus, HeLa, and activated T cell tissues. These cDNAs have been placed in a linear order by mapping within the CEPH mega YAC contigs and/or radiation hybrids. Additionally, a method has been developed for directly hybridizing YACs to the cDNA arrays, quickly identifying and binning cDNAs to particular regions of interest. As a result of this approach, chromosome 5 has been annotated with a number of previously characterized but unmapped genes, novel genes containing well-characterized motifs, and completely novel genes. Many of these cDNAs were found to be derived from low abundance transcripts. Moreover, as a subset of these cDNAs were oligo-dT primed, they are directly comparable to the I.M.A.G.E. consortium ESTs. It was therefore possible to ascertain that 30% of the novel genes in this subset do not overlap with previously described ESTs.

Dr. Jan-Fang Cheng presented data on annotating genomic sequence generated from sequencing projects at the Lawrence Berkeley Laboratory's Human Genome Center. The cluster of interleukin genes on chromosome 5q23 was selected as the primary target for this study because of the high likelihood of using sequence information to identify new functionally related genes. The approach involved using a combination of sequence database searches and gene finding programs to identify the location of potential coding sequences. The computer predictions are then tested biologically. Using these approaches, 680 Kb of sequence in a 1 Mb interval containing 6 known genes, IL4, IL13, IL5, IRF1, CSF2 and IL3 were analyzed. A total of 155 EST hits and 484 GRAIL predicted exons were identified. These putative exons were further grouped into an estimated 26 genes based on sequence matches to genes in other organisms and the fact that some predicted

exons form groups by their locations. Of these 26 predicted transcripts, six were the known genes previously mapped to this region, one was a known gene whose location was previously undetermined, 11 matched with genes in other organisms, five were novel genes identified on the basis of EST hits, and three were predicted genes whose expression was verified biologically using a combination of RT-PCR and Northern analysis.

Dr. C. Fidler presented a transcript map of the ADRB2 - NKSF1 interval which had been generated using the combined techniques of cDNA library screening with radiolabelled YACs and the mapping of chromosome 5-specific ESTs to a YAC contig. Using YACs to screen a fetal brain cDNA library resulted in the identification of five genes; comprising two known genes previously localised to the region, one known gene previously unlocalised and two novel human cDNAs.

Map integration

Using the Whitehead Institute and Genethon maps considerable emphasis was placed on the alignment and integration of the data presented together with that available electronically. The full map is available for down-loading via the chromosome 5 web page (http://genome.wustl.edu/gsc/chrom5/Chrom5_home.html).

Resources

WWW

Chromosome 5 web pages:

Location: Washington University School of Medicine, St. Louis MO USA

Curator: John D. McPherson

URL: http://genome.wustl.edu/gsc/chrom5/Crom5_home.html

Partial Index: This report

- Links to Chromosome 5 data and sites

- Latest version of Chromosome 5 AceDb database (5ace 3.0)

- database to download

- web searchable database

- Integrated Chromosome 5 maps

- to download

- viewable

- Index of chromosome 5 sequence projects

- Mirrored data from UCI (see below)

Location: University of California, Irvine, Irvine CA USA

Curator: The UCI Genome Center has closed; however, all of the data will be maintained at this site as long as possible.

URL: <http://chrom5.hsis.uci.edu>

Partial Index: Chromosome 5 AceDb database version 2.0

- contains all data from the UCI Genome Center

- gif images from 5ace 2.0

- RH raw data

Sequencing Centers:

Washington University School of Medicine

Genome Sequencing Center

URL: <http://genome.wustl.edu/gsc/gschmpg.html>
Current regions: 5p15.2, 5q12-q13, 5q31

Lawrence Berkeley National Laboratory
Human Genome Center
URL: <http://www-hgc.lbl.gov/GenomeHome.html>
Current regions: 5q31

General Clone Resources:

Research Genetics
URL: <http://www.resgen.com>

Genome Systems
URL: <http://www.genomesystems.com>

References

Gyapay G. et al. (1994) The 1993-1994 Genethon human genetic linkage map. *Nature Genet.* 7, 246-339.

Villamar M, Gomendio B, Perera E, Telleria D, Fizames C, SanMillan JL, Weissenbach J, Schott JJ, Dixon M, Moreno F, Hollyoake M (1994) The EUROGEN map of human chromosome 5. *Eur J Hum Genet* 2, 212-213.

Weissenbach J. et al. (1992) A second generation linkage map of the human genome. *Nature* 359, 794-801.

Hudson, T.J., L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista, L. Kruglyak, S.H. Xu, et al. 1995. An STS-based map of the human genome. *Science* 270: 1945-54. With supplementary data from the Whitehead Institute/MIT Center for Genome Research, human genetic mapping project, data release 10 (May 1996).

Schuler, G.D., M.S. Boguski, E.A. Stewart, L.D. Stein, G. Gyapay, K. Rice, R.E. White, P. Rodriguez-Tome, A. Aggarwal, E. Bajorek, et al. 1996. A gene map of the human genome. *Science* 274: 540-546.

Takahashi, E., A. Hitomi, and Y. Nakamura. 1993. A high-resolution cytogenetic map of human chromosome 5: localization of 206 new cosmid markers by direct R-banding fluorescence in situ hybridization. *Genomics* 17: 234-6.

Participants

Dr. Jan-Fang Cheng,
Human Genome Center,
Lawrence Berkeley Laboratory,
1 Cyclotron Road, MS 74-157,

Berkeley, CA 94720.
United States of America.

Tel: +1-510 486 6590
Fax: +1-510 486 6816
E-mail: jcheng@genome.lbl.gov

Dr. G. Clines,
University of Texas Southwestern, Medical Center,
6000 Harry Hines Blvd.
NB10.106, Dallas,
TX 75235-8591.
United States of America.

Tel: +1-214 648 1676
Fax: +1-214 648 1666
E-mail: clines@mcdermott.swmed.edu

Dr. M.J. Dixon
, School of Biological Sciences,
3.239, Stopford Building,
University of Manchester,
Oxford Road, Manchester, M13 9PT
United Kingdom

Tel: +44-161 275 5620
Fax: +44-161 275 5620
E-mail: mdixon@fs2.scg.man.ac.uk

Dr. Nicholas Fairweather,
B.I.R.D.
University of Bath,
Trim Bridge,
Bath.
BA1 1HD.
United Kingdom

Tel: +44-1225 448444
E-mail: prsndf@Bath.ac.uk

Dr. Carrie Fidler,
LRF Molecular Haematology Unit,
Department of Cellular Science,
John Radcliffe Hospital,
Headington,
Oxford. OX3 9DU.
United Kingdom.

Tel: +44-1865 220366
Fax: +44-1865 221778
E-mail: cfidler@hgmp.mrc.ac.uk

Dr. Michelle Le Beau,
University of Chicago,

Section of Hematology/Oncology,
5841 South Maryland Avenue,
MC 2115,
Chicago, IL 60439.
United States of America.

Tel: +1-773 702 0795
Fax: +1-773 702 0963
E-mail: mmlebeau@mcis.bsd.uchicago.edu

Dr. Bronwen Loder,
HUGO SCW Contract,
Department of Pathology,
Cambridge University,
Tennis Court Road,
Cambridge. CB2 1QP.
United Kingdom.

Tel: +44-1223 339398
Fax: +44-1223 339398
E-mail: hugo@mole.bio.cam.ac.uk

Dr. J.D. McPherson,
Department of Genetics,
Genome Sequencing Center,
Washington University School of Medicine,
Box 8501,
4444 Forest Park Blvd.
St. Louis, MO 63108.
United States of America.

Tel: +1-314 286 1848
Fax: +1-314 286 1810
E-mail: jmcphers@watson.wustl.edu

Conover Talbot, Jr.
Genome Database
Johns Hopkins School of Medicine
2024 E. Monument St.
Baltimore, MD 21205-2100.
United States of America.

Tel: +1-410 614 0443
Fax: +1-410 614 0434
E-mail: cct@gdb.org