

Mouse Sequencing Consortium

*For release: Friday, October 6, 2000
12:01 am*

*Contact for the Consortium
Mary Prescott
1-312-397-6604
mprescott@bsmg.com*

*(note to reporters/editors: see last
pages for more contacts)*

Public-Private Consortium to Accelerate Sequencing of Mouse Genome

Results will expedite discovery of human genes

The National Institutes of Health, the Wellcome Trust and three private companies today announced they have formed a consortium to speed up the determination of the DNA sequence of the mouse genome. The Mouse Sequencing Consortium will provide \$58 million over the next six months to decipher the mouse genetic code.

Members of the Mouse Sequencing Consortium (MSC) and their contributions to the effort are SmithKline Beecham (\$6.5 million), the Merck Genome Research Institute (\$6.5 million), Affymetrix, Inc. (\$3.5 million), the Wellcome Trust (\$7.75 million), and six of the National Institutes (\$34 million*), including the National Cancer Institute, the National Human Genome Research Institute, the National Institute on Deafness and Other Communication Disorders, the National Institute of Diabetes and Digestive and Kidney Disease, the National Institute of Neurological Disorders and Stroke, and the National Institute of Mental Health.

MSC funds will support mouse genome sequencing at three DNA sequencing laboratories: the Whitehead Institute for Biomedical Research in Cambridge, Mass., Washington University School of Medicine in St. Louis, and the Sanger Centre in the U.K.

The MSC is another example of an emerging model for supporting large-scale genomics research in which public and private sector entities join forces to produce publicly available data sets that are crucial for basic biomedical research.

Like the efforts of The SNP Consortium (a group of pharmaceutical and technology companies that together with the Wellcome Trust are constructing a map of genetic variations that occur throughout human DNA) and the Merck-funded effort to generate a database of expressed sequence tags (DNA known to match regions of the genome that code for proteins), the MSC is a public-private partnership to generate data that will be freely available for the unrestricted use of biomedical researchers worldwide. Private sector participation in the MSC has been facilitated by the Foundation for the National Institutes of Health, Inc., a non-profit, charitable organization founded to support the NIH in its mission.

* Precise funding levels for the National Institutes of Health are contingent upon final fiscal year 2001 budget appropriations to be passed by the U.S. Congress.

The desire to accelerate mouse genome sequencing builds on the completion in June 2000 of the working draft version of the human DNA sequence. With the working draft of the human genome sequence in hand, scientists in both industry and academia now seek to interpret its meaning. The DNA sequence of the mouse genome will provide an essential tool to identify and study the function of human genes.

Sequencing the mouse genome is now the next major goal of large-scale genomics and the Mouse Sequencing Consortium's effort will expand and accelerate the program to analyze the mouse genome begun by the National Human Genome Research Institute (NHGRI) in September 1999. That program already has generated most of the data for a "fingerprint" map of the mouse genome, including a set of sequences from the ends of cloned genomic DNA fragments, and is doing targeted sequencing of regions of the mouse genome that are of particularly high biological interest. The NHGRI effort also has begun to sequence the mouse genome in its entirety.

Mammals share many basic biological functions such as immune response, regulation of cell division, and development of major organ systems. The gene sequences in mouse and human that encode the proteins to carry out these functions also are shared to a high degree (85% sequence identity). The DNA sequences in the vast regions between genes are much less similar (50% sequence identity or less).

Since only about 5% of the human genome contain genes, sifting through the 3.1 billion DNA letters to find genes is an extremely challenging task. But, by comparing human and mouse genome sequences, the regions of high similarity are readily apparent and immediately identify protein coding regions and regulatory sequences. Thus, the mouse genome sequence will provide a powerful tool to interpret the newly available human genome sequence.

In addition to its use to aid the interpretation of the human genome, the mouse genome sequence also will increase the ability of scientists to use the mouse as a model system to study and understand human disease, and to develop and test new treatments in ways that can not easily be done with humans.

The genome of the mouse is the same size as that of the human, about 3.1 billion base pairs. As recommended by scientists studying the mouse, the genome sequencing effort will use a strain of mouse known as C57BL6/J, commonly called "Black 6." The sequencing strategy that will be used takes advantage of the best features of the map-based shotgun strategy used by the public sequencing consortium to produce the human sequence and the whole genome shotgun strategy used by the private sector effort that also produced a version of the human genome sequence in the past year. The melding of these two strategies promises to produce a high quality genome sequence more quickly than either strategy could alone.

The MSC's program will, by the end of February 2001, bring the overall depth of coverage of the mouse genome to 2.5X to 3X. This is the level of coverage at which shotgun genomic sequence first becomes useful to the typical scientist, with about 93 to 95 percent of the

sequence of the mouse genome being available albeit in small, unordered fragments. Subsequently, the mouse genome sequencing effort will generate the complete sequence coverage and assemble the entire sequence into a "finished," highly accurate form.

The data release practices of the MSC will continue the international Human Genome Project's sequencing program's objective of making sequence data available to the research community as soon as possible for free, unfettered use. In fact, the incorporation of the whole genome shotgun sequencing component has led to adoption of a new, even more rapid data release policy whereby the actual raw data (that is, individual DNA sequence traces, about 500 bases long, taken directly from the automated instruments) will be deposited regularly in a newly-established public database operated by the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/) and a sister database operated by the European Bioinformatics Institute (EBI, www.ebi.ac.uk). These individual DNA sequences will be assembled into larger assemblies as soon as sufficient coverage is attained, which will be at about the point where working draft quality coverage of the genome is reached.

###

For more information:

Mouse Sequencing Consortium Members

SmithKline Beecham

Merck Genome Research Institute

Affymetrix, Inc.

National Cancer Institute

National Human Genome Research Institute

National Institute on Deafness and Other
Communication Disorders

National Institute of Diabetes and Digestive
and Kidney Diseases

National Institute of Mental Health

National Institute of Neurological Disorders
and Stroke

Wellcome Trust

Media Contacts

Graeme P. Holland
44-12-7964-4269
Graeme_P_Holland@sbphrd.com

Rick Koenig
1-610-270-5546
Rick_M_Koenig@sbphrd.com

Kathryn Munoz
1-908-423-6492
kathryn_munoz@merck.com

Anne Bowdidge
1-408-731-5925
anne_bowdidge@affymetrix.com

NCI Press Office
301-496-6641

Cathy Yarbrough
1-301-594-0954
cyarbrou@mail.nih.gov

Marin Allen
1-301-496-7243
marin_allen@nih.gov

Joan Chamberlain
1-301-496-3583
joan_chamberlain@nih.gov

Marilyn Weeks
1-301-443-4536
mweeks@nih.gov

Margo Warren
1-301-496-5751
mw76v@nih.gov

Noorece Ahmed
44-20-7611-8540
n.ahmed@wellcome.ac.uk

Genome Sequencing Centers

Whitehead Institute for Biomedical Research

Washington University School of Medicine

Sanger Centre

Foundation for the National Institutes of Health, Inc.

Other Contacts

Media Contacts

Seema Kumar
1-617-258-6153
kumar@wi.mit.edu

Joni Westerhouse
1-314-286-0120
joniw@medicine.wustl.edu

Don Powell
44-12-2349-4956
don@sanger.ac.uk

Constance U. Battle, MD
1-301-402-5311
cubattle@fnih.org

Arthur Holden
1-773-867-2990
aholden@earthlink.net