Table S1: Groups participating in the ENCODE Pilot phase.

Initial ENCODE Pilot Phase Participants

Research Group	Institution	Research Goals
Ian Dunham	Wellcome Trust Sanger Institute	Map origins of replication, DNA methylation, chromatin modifications, transcription factor binding sites, primarily with ChIP-chip assays using spotted DNA microarrays.
Anindya Dutta	University of Virginia	Identify early and late origins of replication, sites of replication termination and pause sites for replication forks. Replication products mapped by hybridization to Affymetrix microarrays.
Thomas Gingeras	Affymetrix, Inc.	Map RNA transcripts, binding sites for transcription factors and chromosomal proteins using Affymetrix microarrays and ChIP-chip assays.
Roderic Guigó	Municipal Institute of Medical Research	Identify all protein-coding genes. Combine computational prediction with experimental RT/PCR confirmation of gene models.
Richard Myers	Stanford University	Identify promoters and enhancers with transfection of reporter constructs into cell lines. Identify transcription factor binding sites and chromatin modifications with ChIP-chip assays using Nimblegen and Agilent arrays. Identify conserved domains with comparative genomic data. Test the function of conserved domains by mapping polymorphisms to these domains and assaying for the effect of these polymorphisms in reporter assays for enhancers in transfected cells.
Bing Ren	Ludwig Institute for Cancer Research	Identify promoters, enhancers, repressors/silencers using ChIP-chip assays and mapping on spotted DNA microarrays.
Michael Snyder	Yale University	Map RNA transcripts and binding sites for transcription factors and chromosomal proteins using DNA microarrays and ChIP-chip assays. Comparison of Affymetrix, NimbleGen and spotted DNA arrays platforms.
George Stamatoyannopoulos	University of Washington	Map DNase I hypersensitive sites using quantitative, real time PCR.

Additional ENCODE Pilot Phase Participants

Research Group	Institution	Research Goals
Andy Baxevanis	National Human Genome Research Institute	Develop an ENCODE data portal for non-sequence based data including coordinated data deposition and dissemination.
Kerstin Lindblad-Toh/ Michelle Clamp	Broad Institute	Develop methodologies, algorithms and software to generate regional alignments of multiple genomes in the ENCODE regions.
Greg Crawford/ Francis Collins	National Human Genome Research Institute	Identify DNase hypersensitive sites; develop high-throughput Massively Parallel Signature Sequencing (MPSS) assay for DNase hypersensitive sites.
Pieter De Jong	Children's Hospital Oakland Research Institute	Create clone resources to support comparative sequencing.
Eric Green	NIH Intramural Sequencing Center/ National Human Genome Research Institute	Isolate BAC clones for ENCODE regions in multiple organisms; generate multispecies comparative genome sequence data for these ENCODE regions; develop computational tools for analysis of comparative genome sequences.
Ross Hardison	Pennsylvania State University	Develop tools to analyze comparative genomic sequences and integrate functional data with the genome sequence.
David Haussler	University of California, Santa Cruz	Develop ENCODE-specific views of the human genome using the Santa Cruz UCSC Browser; develop tools to analyze comparative genomic sequences and integrate functional data with the genome sequence.
Steven Jones	British Columbia Cancer Agency Genome Sciences Centre	Generate whole genome data on gene expression; develop tools to identify regulatory elements from co-expressed genes.
Marco Marra	British Columbia Cancer Agency Genome Sciences Centre	Generate fingerprint maps and tiling paths for BACs isolated from the ENCODE regions in different species; identify alternatively spliced transcripts for genes in the ENCODE regions.
Webb Miller	Pennsylvania State University	Develop tools to analyze comparative genomic sequences and integrate functional data with the genome sequence.
Steve Salzberg	The Institute for Genomic Research	Develop computational tools to analyze comparative genomic Sequences, to find genes and to assemble genomes.
Greg Schuler	National Center for Biotechnology Information (NCBI), National Library of Medicine	Coordinate ENCODE comparative genomic sequence data with NCBI.

Table S2: Groups participating in the ENCODE Technology Development phase.

ENCODE Technology Development Phase Participants

Research Group	Institution	Research Goals
Job Dekker[1]	University of Massachusetts Medical School	Develop PCR strategy to identify regions in chromosomes that interact through protein complex binding using the Chromosome Conformation Capture (3C) technology.
Xiang-Dong Fu[1]	University of California, San Diego	Improve sensitivity and specificity of the ChIP-chip technology using single stranded oligonucleotide microarrays and DASL (DNA Annealing Selection and Ligation) technology.
Roland Green[1]	NimbleGen Systems, Inc.	Test ability of NimbleGen's Maskless Array Synthesis technology to map transcription factor binding sites and first exon/promoter identification in ChIP and microarray assays.
Robert Kingston[1]	Massachusetts General Hospital	Develop high-throughput methods for mapping chemical and enzymatic DNA cleavage sites in chromatin at nucleotide resolution.
Mark McCormick[1]	NimbleGen Systems, Inc.	Develop "exon-linkage assay" to study alternative splicing using NimbleGen's oligonucleotide array platform.
Zhiping Weng[1]	Boston University	Develop computational methods to identify cis-regulatory elements in alternative promoters and confirm these elements by competitive PCR and reporter-construct assays in transfected cells.
Joseph Ecker[2]	Salk Institute for Biological Studies	Test multiple methylation detection methods (bisulfite treatment, restriction digest with methylation-sensitive enzymes, anti-methylcytosine antibodies, binding of proteins to methylated DNA) using whole genome microarrays.
Vishwanath Iyer[2]	University of Texas, Austin	Develop methods for Sequence Tag Analysis of Genomic Enrichment (STAGE) that combines chromatin immunoprecipitation and SAGE and Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) to identify regulatory elements in chromatin.
Madaiah Puttaraju[2]	Intronn, Inc	Develop methods to find splice sites using pre-trans splicing molecules (Spliceosome Mediated RNA Trans-splicing (SMaRT) ExonFinder).

Yijun Ruan[2]	Genome Institute of Singapore	Develop Gene Identification Signature (GIS) analysis to generate SAGE data with linked 5' and 3' mRNA information; extend this method to analysis ChIP DNA fragments.
Scott Tenenbaum[2]	The Research Foundation of SUNY, Albany	Develop methods to detect protein binding sites in mRNA by ribonucleoprotein immunoprecipitation combined with microarray hybridization (RIP-chip).
Thomas Tullius[2]	Boston University	Develop a library of hydroxyl radical cleavage patterns of random DNA and use to develop computational predictions of DNA cleavage patterns and correlation with DNA binding sites.

[1]Projects were initiated September 2003
[2]Projects were initiated September 2004

Table S3: Anticipated Comparative Sequencing Datasets for the ENCODE Target Regions

Name	Latin Name	Sequence Quality	Source
Armadillo (nine-banded)	*Dasypus novemcinctus*	Comparative grade finished*	NIH Intramural Sequencing Center
Baboon (olive)	*Papio cynocephalus anubis*	Comparative grade finished	NIH Intramural Sequencing Center
Bat (greater horseshoe)	*Rhinolophus ferrumequinum*	Comparative grade finished	NIH Intramural Sequencing Center
Cat	*Felis catus*	Comparative grade finished	NIH Intramural Sequencing Center
Dusky titi	*Callicebus moloch*	Comparative grade finished	NIH Intramural Sequencing Center
Elephant (African)	*Loxodonta africana*	Comparative grade finished	NIH Intramural Sequencing Center
Galago (small-eared)	*Otolemur garnetti*	Comparative grade finished	NIH Intramural Sequencing Center
Guinea Pig	*Cavia porcellus*	Comparative grade finished	NIH Intramural Sequencing Center
Hedgehog (middle-African)	*Atelerix albiventris*	Comparative grade finished	NIH Intramural Sequencing Center
Lemur (gray mouse)	*Microcebus murinus*	Comparative grade finished	NIH Intramural Sequencing Center
Marmoset (white-tufted ear)	*Callithrix jacchus*	Comparative grade finished	NIH Intramural Sequencing Center
Monkey (colobus)	*Colobus guereza*	Comparative grade finished	NIH Intramural Sequencing Center
Monkey (owl)	*Aotus nancymaae*	Comparative grade finished	NIH Intramural Sequencing Center
Platypus (duck-billed)	*Ornithorhynchus anatinus*	Comparative grade finished	NIH Intramural Sequencing Center
Rabbit	*Oryctolagus cuniculus*	Comparative grade finished	NIH Intramural Sequencing Center
Shrew (European common)	*Sorex araneus*	Comparative grade finished	NIH Intramural Sequencing Center
Tenrec (lesser hedgehog)	*Echinops telfairi*	Comparative grade finished	NIH Intramural Sequencing Center
Bovine	*Bos taurus*	High-quality finished**	Baylor College of Medicine Human Genome Sequencing Center
Chimpanzee	*Pan troglodytes*	High-quality finished	Washington University Genome Sequencing Center, Broad Institute/MIT Center for Genomic Research

Dog	*Canis familiaris*	High-quality finished	Broad Institute/MIT Center for Genomic Research
Frog	*Xenopus tropicalis*	High-quality finished	US Department of Energy Joint Genome Institute
Macaque	*Macaca mulatta*	High-quality finished	Baylor College of Medicine Human Genome Sequencing Center, Washington University Genome Sequencing Center, J. Craig Venter Joint Technology Center
Mouse	*Mus musculus*	High-quality finished	Washington University Genome Sequencing Center, Broad Institute/MIT Center for Genomic Research, Wellcome Trust Sanger Institute
Rat	*Rattus norvegicus*	High-quality finished	Baylor College of Medicine Human Genome Sequencing Center
Zebrafish	*Danio rerio*	High-quality finished	Wellcome Trust Sanger Institute
Chicken	*Gallus gallus*	High-coverage whole genome shotgun***	Washington University Genome Sequencing Center
Opossum	*Monodelphis domestica*	High-coverage whole genome shotgun	Broad Institute/MIT Center for Genomic Research
Orangutan	*Pongo pygmaeus*	High-coverage whole genome shotgun	To be determined

* Comparative grade finished sequencing involves shotgun sequencing to 8X-10X coverage with additional manual refinement to order and orient contigs. The product is at an intermediate level between purely shotgun and high-quality finished sequence.

** High-quality finished reflects highly accurate and contiguous sequence, with a best-faith effort used to resolve all difficult regions.

*** Sequence orthologous to the human ENCODE targets generated from whole genome efforts in other organisms will be incorporated into the ENCODE dataset where available. The ultimate product of these efforts may vary in terms of depth of shotgun coverage.