Identification of Transcribed Sequences: Functional and Expression Analysis

November 16-19, 1997 Asilomar Conference Center, Monterey, CA

Meeting Information

Topics to be discussed include but are not limited to: mammalian gene and genome organization as determined from the construction of transcriptional maps and genomic sequencing; expression analysis of novel mammalian genes; analysis of genomic sequence, including gene and regulatory sequence prediction and verification, and annotation for public databases; expression and mutation analysis, and comparative mapping, in model organisms (e.g. yeast, C.elegans, drosophila, zebrafish, pufferfish, chicken, mouse, rat); construction and analysis of transgenic organisms; novel approaches for functional analysis of transcribed sequences; construction of full length and 5' specific cDNA libraries; and database construction, management and use in expression and functional analyis.

IWITS 97 Sponsors

- Amgen
- Clonetech
- Digital Gene Technologies

The format of the workshop is designed to foster general discussion with the majority of attendees presenting 15 minute talks. Attendance is limited and dependent upon acceptance of an abstract.

Organizers

Katheleen Gardiner
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80206
303-333-4515

Chris Lau
UCSF
415/476-8839

Richard J. Mural
Oak Ridge National Laboratory
P.O. Box 2009
Oak Ridge, TN 37831- 8077
423-576-2938

Provisional Agenda

Abstracts

Provisional Schedule

7th International Workshop on the Identification of Transcribed Sequences

November 17, 1997

	Session 1. Chair: Nguyen	
8:30	Greetings, Opening Remarks	
8:40	Werner	Prediction of muscle-specific gene expression: the actin promoter model
9:00	Soares	Development of libraries enriched for full-length cDNAs. A progress report
9:20	Wolfe	Whole-genome duplication during yeast evolution and its (ir)relevance to other organisms
9:40	Gardiner	Gene discovery in AT-rich regions of human chromosome 21 and correlation of base composition with compaction in Fugu rubripes homologous genes
10:00	Break	
10:20	Postlethwait	A linkage map of zebrafish transcribed sequences and the evolution of the vertebrate genome
10:40	Ross MacDonald	S. cerevisiae, lacZ., lox, in vivo functional analysis of the S. cerevisiae genome, the transposon insertion project, the genome deletion project
11:00	Brooks	Database assisted large scale polymorphism finding and exploitation
11:20	Lewis	A report on the Drosophila EST project
11:40	Overton	EPODB: A bioinformatics system for the analysis of gene expression during erythropoiesis
12:00 - 1:00	Lunch	
	Session 2. Chair: McClelland	
2:30	Fontes	Utilisation of YACs as expression vectors, to create or correct pathological phenotypes.
2:50	Lin	Functional studies of coding and non-coding sequences of zebrafish genome
3:10	Hasel	An automated digital technology for analyzing expression of nearly all genes
3:30	Break	
4:00	Leung	mRNA architecture and cascade of gene expression
4:20	Poustka	Evolutionary history of genes from the G6PD region of human Xq28: Insights from the analysis of human, Fugu and amphioxus homologues
4:40	Guigo	Attaching functional annotation to predicted genes in genomic sequences
6:00 - 7:00	dinner	

7:30 - 9:00	**Poster Symposium Chair: Yaspo**	
	Udar	Comparative analysis and genomic structure of the Ataxia telangiectasia gene in human and pufferfish and characterization of some founder mutations.
	Lubyova	Isolation and characterization of a novel human gene, hsUFD2, located in the neuroblastoma critical region on chromosome 1p36.3.
	Smith	Analysis of the chicken genome by means of a sequence sampling approach.
	Wright	Structure/function analysis of vertebrate Na+/H+ exchanger genes and their role in control of cell volume.
	Onyango	A bacterial artificial chromosome expression system for scanning large DNA segments for functional elements: application in neuroblastoma.
	Wiemann	Generation and sequencing of full length cDNAs in the course of the German Genome Project.
9:00 - 11:00	Cash Bar	

November 18, 1997

	Session 3. Chair: Poustka	
8:30	McClelland	GeneUP: A program to select short PCR primer pairs that occur in multiple members of sequence lists.
8:50	Nguyen	Quantitative anslysis of differentially expressed cDNA between normal and deficient thymus lead to the identification of novel functional genes
9:10	Korn	X chromosome cDNA library: Generation, evaluation and application in transcriptional mapping
9:30	Ashburner	Gene function - a model for the representation of gene function in a database - FlyBase
9:50	Break	
10:20	Nadeau	Phenotypic dissection of complex traits in mice: risk factors, expression profiles, physiological and developmental pathways, and models for human genetic diseases
10:40	Rynditch	Chromosome 3q breakpoints in leukemia: complexities in alternative processing and intergenic splicing
11:00	Pesole	Structural and compositional features of untranslated regions of eukaryotic mRNAs
11:20	Jansen	Reverse genetics by chemical mutagenesis in C. elegans
11:40	Peterfy	SEX trap: A novel method for the identification of signal sequence-containing genes from cytokine clusters
12:00 - 1:00	Lunch	
	Session 4. Chair: ?	
2:30	Kohara	Expression pattern map of the C. elegans genome
2:50	Ross	Use of cDNA microarrays for analysis of gene expression for thousands of genes simultaneously
		FDD, a high-throughput message display system, for the functional interpretation of

3:10	Ito	yeast and mammalian genomes
3:30	Mural	Automated annotation of genomic DNA sequences: The genome channel
3:50	Break	
4:20	Den Dunnen	Gene identification in sequenced DNA: database searches, gene predictions and verification by RT-PCR, Northern blots and exon trapping
4:40	Leanna	Development and use of the reverse two-hybrid system to characterize interactions between c-Rel and its inhibitor, IkBa
5:00	Ringwald	Gene expression resource for the laboratory mouse
6:00 - 7:00	dinner	
7:30 - 9:00	**Poster Symposium Chair: Soares**	
	Rao	Growth failure in idiopathic short stature and Turner syndrome is caused by haploinsufficiency of a pseudoautosomal homeobox gene
	Lundeberg	Analysis of differentially expressed genes using an solid-phase RDA approach
	Jokadze	Profile the expression of hundreds of human genes
	Woloschak	Identification of consensus elements in 3' UTRs.
	Slavov	RNA editase genes from Fugu rubripes
	Zhang	In situ protein blotting: an efficient method to detect protein-protein interactions on tissue sections
9:00 - 11:00	Cash Bar	

November 19, 1997

	Session 5. Chair: Hasel	
8:30	Freeman	Expression mapping of mouse genes
8:50	Lord	Comparative mapping in the search for the ovine Booroola mutation
9:10	Yaspo	Genomic sequence analysis of a gene dense region on chromosome 21q22.3, and comparison with the existing transcript map: the emerging gene organization.
9:30	Kingsmore	Identification of ANF as a stroke susceptibility gene by positional expression cloning using quantitative expression analysis
9:50	break	
10:20	Zhu	Linking human proteins using two-hybrid technology
10:40	Break	
11:00	Lau	The mouse Sry interactive proteins are differentially expressed in adult and fetal tissues
11:20	Closing remarks/adjourn	

Identification of Transcribed Sequences: Functional and Expression Analysis
Meeting Abstracts (November 1997)

Utilisation of YACs as expression vectors to create or correct pathological phenotypes

E.Passage, Z.Assouline, D.Sabéran-Djoneidi, M.Fontés
INSERM U406, Génétique Médicale et Dévelopement, Fac. de Médecine, 27 BD. J.Moulin, 13358 Marseille Cedex 5, France

A major goal in human genetics is to create good animal models for human inherited pathologies to: 1 - Understand the physiopathology of the disorder. 2 - Test new therapeutical approaches. Standard transgenic approaches have been mainly proposed, but they result in random integration of multiple copies of transgenes. The position effects related to the integration site often lead to low levels of gene expression, and aberrant patterns of expression. Moreover, the large copies number of tandemly repeated sequences is frequently unstable. We have thus developed, in collaboartion with C.Huxley (London), an alternative approach, using large genomic fragments (as YACs) as vectors. The large size of YACs (several hundreds of Kb) usually transfer all elements required for faithful regulation of gene expression, in quantitative as well as in qualitative terms. We have applied this technology to create an animal model of the Charcot-Marie-Tooth disease type 1A which is the most frequent inherited peripheral neuropathy in human. This disorder is caused by a 1.5 Mb duplication, which includes the gene coding the synthesis of the myelin protein PMP22. We have injected a human YAC of 560 Kb, containing PMP22 and flanking elements, in the murine oocytes. We have obtained 5 lines, which have integrated from 1 to 7 copies of the YACs. The human PMP22 gene expression has been determined, and we have demonstrated that gene expression is proportional to the copy number of the YAC integrated in the murine genome. Mice with 1/2 copies have no phenotypes. Mouse with 4 copies have a demyelinating neuropathy comparable to CMT1A. Mouse with 7 copies have a very severe peripheral neuropathy. This technique has thus proven to be extremely powerful to create a pathological phenotype. Moreover, we think that if the technique is able to create a phenotype, it can also be used to correct a phenotype. In that way, we will present data of YACs expression in human cells, and particularly the expression of CFTR gene in normal and CF cells.

Michel Fontés
INSERM U406- Fac de Médecine
27 Bd.J.Moulin
13358 MARSEILLE CEDEX5 FRANCE
telephone: 33 4 91 25 71 59
fax: 33 4 91 80 43 19
email: MICHEL.FONTES@MEDECINE.UNIV-MRS.FR

Presentation format: Platform

Isolation and characterization of a novel human gene, hsUFD2, located in the neuroblastoma critical region on chromosome 1p36.3

Barbora Lubyova, Ute Willhoeft, Patrick Onyango, Judith A. Lummerstorfer and Andreas Weith
Institute of Molecular Pathology (IMP), Vienna, Austria

Chromosome 1p36 region consistently displays allelic deletions in a variety of cancer, e.g. melanoma, neuroblastoma, breast cancer, or hepatoma. We previously defined a chromosomal segment, 1p36.31 - p36.32, whose integrity is essential for tumor suppression in neuroblastoma. Applying a positional cloning strategy we have identified and cloned a novel human gene located in the neuroblastoma critical region, in the immediate vicinity of a balanced translocation breakpoint found in a primitive neuroectodermal tumor (PNET). Structural analysis of this novel gene by long range sequencing revealed 28 exons and a genomic coverage of the transcribed region of more than 200 kb. Northern blot analysis displayed a fairly wide expression pattern with predominant transcript length of 6.5 kb and a putative alternative splice product of 6.2 kb in most tissues. There is an additional 2.5 kb transcript present in skeletal muscle. The cDNA contains an open reading frame of 3522 nucleotides, encoding a putative 100 kD protein. The predicted protein sequence contained no obvious functional sequence motifs; however, database comparisons revealed a very high homology to the yeast UFD2 gene (Ubiquitin fusion degradation protein). UFD2 represents a factor involved in ubiquitin-associated protein degradation pathways. We designate the chromosome 1p36.3 located gene hsUFD2 owing to the homologies found. Initial screening attempts for mutations have revealed a gross rearrangement of this gene in at least one primary hepatoma tumor. Independent of the structural analyses we have recently started to analyze the function of hsUFD2 in order to gain insight into its possible role in malignant transformation. Functional assays employing transfection analyses in neuroblastoma tumor cell lines indicate that expression of this gene is incompatible with soft agar growth of tumor cells. Immunocytochemical and phenotypic analyses of transfected neuroblastoma clones are currently in progress.

Barbora Lubyova
Institute of Molecular Pathology (I.M.P.)
Dr. Bohr-gasse 7
Vienna, Austria A-1030
telephone: +431-79730-423
fax: +431-7987153
email: Lubyova@aimp.una.ac.at

Presentation format: Poster

Analysis of the chicken genome by means of a sequence sampling approach

J. Smith, R.I. Paton, C. Bruley, A.S. Law, and D.W. Burt

The chicken genome comprises eight pairs of large autosomal 'macrochromosomes', Z and W sex chromosomes and thirty pairs of small 'microchromosomes'. Work has been done which indicates that there is a higher concentration of CpG islands on the microchromosomes than is present on the macrochromosomes. Knowing that 60-70% of chicken genes are associated with CpG islands, this may imply that the microchromosomes (accounting for only 25% of the genome) are more gene dense than the macrochromosomes .

In order to test this theory, a 'sequence sampling' approach has been undertaken. To date, ten cosmids which have been physically assigned to chicken chromosomes by FISH, have been subcloned and a representative portion of each cosmid sequenced (70%). Five cosmids are known to map to macrochromosomes and five to microchromosomes. Each sequence is searched against blastn, blastp, dbest and dbsts databases as well as the Fugu rubribes database. CpG analysis is also carried out on each sequence. This is carried out automatically using a system based on the Pregap part of the Staden processing programme. for each cosmid, the presence of genes, gene homologies, CR1 repeats and CpG islands is noted. So far there is no indication that microchromosomes are more gene dense than their macrochromosomal counterparts. As a means of randomly finding genes within a particular region of DNA, sequence sampling is proving itself to be a powerful tool.

Dr. Jacqueline Smith
Roslin Institute
Roslin, Midlothian EH25 9PS
Scotland
telephone: 44 (0) 131 527 4200
fax: 44 (0) 131 440 0434
email: Jacqueline.Smith@bbsrc.ac.uk

Presentation format: Poster

S. cerevisiae , lacZ., lox , in vivo Functional analysis of the S. cerevisiae genome: The Transposon Insertion Project, The Genome Deletion Project

Petra Ross-Macdonald, S. Agarwal, R. Bangham, H. Liao, A. Sheehan, D. Symonaitis, L. Umansky, G.S. Roeder, and M. Snyder
Yale Genome Analysis Center, Department of Biology, Yale University, New Haven, CT 06520.

We are involved in two large-scale S. cerevisiae genome projects. The first applies a random transposon insertion/tagging approach; the second will systematically create deletions of each yeast ORF.

The Transposon Insertion Project
(http://ycmi.med.yale.edu/YGAC/home.html): We intend to construct a reporter gene fusion and an epitope-tagged version of the protein for every gene in the yeast genome. This will allow us to analyze gene expression throughout the life cycle, the subcellular location of the encoded protein, and the phenotypic effect of disrupting the gene. A novel transposon system (PNAS 94: 190, 1997) has been used to create a bank of yeast strains, each with a transposon inserted at a random genomic location. This insertion allows us to analyze expression of yeast ORFs via in-frame fusions to lacZ. Using the Cre/lox system, we can modify the transposon in vivo to derive an in-frame element that leaves a 93 amino-acid epitope tag in the product of the mutagenized gene. We have found that in strains carrying insertions in essential genes, the smaller tag often does not cause lethality. Such strains should provide a rich source of conditional or hypermorphic alleles, a valuable genetic tool.

The full-length, tagged proteins are also excellent substrates for immunolocalization. Thus far, we have identified

6500 strains with vegetatively-expressed fusions, and examined the subcellular localization of epitope-tagged proteins in 2112 strains. Sixty strains have tagged proteins that localize to the nucleus; we are extending this work by immunolocalization on spread chromosomes. We have also identified tagged proteins localizing to more unusual sites, such as the bud neck, vacuolar rim and spindle apparatus.

Sequence analysis shows that in addition to finding fusions in recognized ORFs, some highly-expressed fusions occur in large ORFs that were not annotated by the systematic sequencing project, while numerous insertions correspond to fusions to smaller unn amed ORFs. Of the nuclear-localizing tagged proteins, about 60% correspond to insertions in uncharacterized ORFs; many of the remainder are in known chromatin-associated proteins or transcription factors.

The Genome Deletion Project
(http://sequence-www.stanford.edu/group/yeast_deletion_project/deletion.html): In a collaborative effort involving the USA, Canada and Europe, whole-gene deletions are being constructed for each annotated ORF in the yeast genom e. Over the next two years, our laboratory will construct 600 deletion strains. Each deletion construct bears a unique tag; hence parallel analysis of all c. 6000 strains generated will be possible, using hybridization to an Affymetrix chip carring the tag sequences. This will allow both specific target screens and studies on the enviromental interactions of all genes.

Petra Ross-Macdonald
Dept. of Biology, KBT 912
Yale University
PO Box 208103
New Haven, CT 06520-8103
telephone: (203) 432 9949
fax: (203) 432 6161
email: petra@pantheon.yale.edu
url: http://ycmi.med.yale.edu/YGAC/home.html

Presentation format: Platform

GeneUP: A program to select short PCR primer pairs that occur in multiple members of sequence lists

Graziano Pesole1, S Liuni2, G. Grillo3, Pierre Belichard4, Thomas Trenkle4, and Michael McClelland4

1 Dipartimento di Biologia D.B.A.F. Universita' della Basilicata, via Anzio 10, 85100 Potenza Italy 2 Liuni S - Centro di Studio sui Mitocondri e Metabolismo Energetico, C.N.R., via Orabona, 4, 70126 Bari, Italy 3 Grillo G. - Dipartimento di Biochimica e Biologia Molecolare, Universita' di Bari, via Orabona, 70126 Bari, Italy 4 Sidney Kimmel Cancer Center, 3099 Science Park Road, San Diego, CA 92121.

A computer program is described that selects a small set of short primer pairs for PCR to sample all the sequences in a list of mRNAs of interest. Such primer pairs have previously been shown to increase the probability of sampling the mRNAs of interest using RNA fingerprinting. The program selects pairs of primers that have the following properties: [1] each primer pair samples more than one sequence in the list, [2] a small set of primer pairs samples all, or nearly all, of the sequences in the list, [3] the primers have a fixed range of G+C content, [4] primer pairs are excluded that generate simulated PCR products of the same size from a number of

sequences in the list, and [5] primers can be excluded that occur in other lists of sequences. In the examples presented, the primers are confined to 50-90% G+C content and primers are excluded if they occur in the hyper-abundant ribosomal RNAs, mitochondrial RNA, or dispersed transcribed repeats. Pairs of primers of eight or nine bases in length that fit such criteria are generated using four lists; 65 human cDNAs associated with DNA repair; 60 mRNAs associated with apoptosis; 44 members of the human nuclear receptor gene family; 113 members of the G-protein coupled receptor gene family. Applications to much longer lists of mRNAs will be discussed.

Michael McClelland
Sidney Kimmel Cancer Center
3099 Science Park Road
San Diego, CA 92121
telephone: 619 450 5990 ext 280
fax: 619 550 3998
email: mmcclelland@skcc.org
url: www.skcc.org
http://www.skcc.org/skcc_staff/mc_clelland/mc_clelland.html

Presentation format: Platform

Prediction of muscle-specific gene expression: The actin promoter model

Thomas Werner

Due to the enormous amount of new genomic sequences it is mandatory to preselect candidate sequences by computerized analysis prior to experimental functional analysis. This includes prediction of exons and introns as well as the identification of potential regulatory regions which usually encompass multiple regulatory elements that exert their regulatory function only within the correct context. Last year, we reported our approach to this problem and presented sucessful identification of a new LTR as an example. We have now extended our work aiming at the prediction of inherent tissue and/or cell specificity of such regions. Actins comprise one of the most commonly expressed gene families in mammalian tissues. Yet there are specialized actin genes which are either preferentially or exclusively expressed in all or only subsets of muscle cells. These expression patterns are mostly controlled at the level of transcription as is known from Jim Fickett´s work (and his excellent web-site) about muscle-specific gene expression. Therefore, the muscle-specificity of particular actin genes is most likely encoded in their promoter sequences although the most prominent muscle-specific transcription factors MEF2 and MyoD are apparently not crucial in this case although present in some of these promoters. Here, we present a pilot study focusing on the specific recognition of muscle-specific actin promoters. We developed a muscle-actin promoter model starting from a general analysis of the correlation of transcription factor binding sites (TF-sites) with these promoters and identified candidates for crucial TF-sites. Our model consists of 6 different elements and was developed on a training set of 11 sequences. This training set was already to heterogeneous in sequence to allow identification by FASTA analysis. It is muscle-actin specific and does not recognize most of the other muscle-specific promoters indicating that there are several independent ways to achieve muscle-specificity of a promoter. We analyzed more than 150 million bp from GenBank with the actin model and retrieved a total of 63 matches, 34 of which were true muscle-actin matches (54% true positives, there were only 10 false negatives). This demonstrates that specific promoter recognition against a vast background of anonymous sequences is pricipally possible.

Thomas Werner
GSF-National Research Center for Environment and Health
AG BIODV / Institute of Mammalian Genetics
Ingolstaedter Landstrasse 1
Neuherberg, Bavaria D-85758
Germany
telephone: +89-3187-4050
fax: +89-3187-4400
email: werner@gsf.de
url: http://www.gsf.de/biodvThomas WERNER

Presentation format: Platform

Whole-genome duplication during yeast evolution and its (ir)relevance to other organisms

Kenneth H. Wolfe, Cathal Seoighe, Denis C. Shields, Colin A.M. Semple, Lucy Skrabanek
Department of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland

Many pairs of duplicated genes in yeast (*Saccharomyces cerevisiae*) are located within pairs of larger duplicated chromosomal regions, where a set of unrelated genes on one chromosome has a set of paralogs on another chromosome, in conserved map order and transcriptional orientation. About 50% of the yeast genome can be mapped into non-overlapping pairs of duplicated regions like this. These regions appear to be relics of a whole-genome duplication (i.e., a tetraploid stage) during yeast's evolution. Ab out 750 yeast genes (13% of its genome) are members of paralogous pairs identifiably resulting from this duplication, which we estimate occurred about 10^8 years ago. Some of the paralog pairs now have slightly divergent functions or are regulated differ ently. The genome duplication has the consequence that there is a two-to-one correspondence between many yeast genes and their orthologs in other species.

Susumu Ohno suggested in 1970 that tetraploidy played a role in increasing the gene number in vertebrates following their divergence from tunicates. Some large duplicated regions consistent with Ohno's model have recently been mapped in mammals so the st ructure of the yeast genome may provide a model relevant to mammals, although Ohno' s model remains far from proven. The genome of *Caenorhabditis elegans*, on the other hand, shows no sign of large duplicated chromosomal regions but instead has many tandemly repeated duplicate genes (which are virtually absent from yeast).

Gene order evolution in yeast seems to have occurred almost entirely by deletion and reciprocal translocation, with very little transposition of genes. We used computer simulations and analytical methods to estimate that the fraction of genes retained in duplicate after tetraploidy was about 8%, and that the number of "illegitimate" reciprocal translocations which broke up the original duplicated chromosomes into smaller duplicated genomic blocks was about 75. If vertebrate genomes have undergone two rounds of tetraploidy (as often proposed) and similarly small fractions of the genome were retained in duplicate after each round, it may be not be possible either to prove or disprove Ohno's hypothesis for mammals without a near-complete set of mapped and sequenced genes.

Dr. Ken Wolfe
University of Dublin, Trinity College

Department of Genetics
University of Dublin, Trinity College
Dublin 2 Ireland
telephone: +353-1-608-1253
fax: +353-1-679-8558
email: khwolfe@tcd.ie
url: http://acer.gen.tcd.ie/~khwolfe

Presentation format: Platform

Simulation of genetic regulatory circuits

The phage lambda lysis-lysogeny decision is a prototypic genetic switch between alternative phenotypes. Stochastic processes in the gene expression reactions produce an erratic time pattern of signal protein production in individual bacterial cells leading to a wide diversity of signal concentrations across a cell population at any instant. Molecular-level simulations of the lysis-lysogeny decision circuit show that the circuit exploits inherent randomness in these gene expression reaction rates as an essential part of its mechanism. The "noise" in the controlling signals provides the random element that causes the infected population to develop differentially and partition between lytic and lysogenic phenotypes. These phage lambda modeling results show that quantitative simulation of the genetic subsystems that control regulatory bifurcation points and environmentally-induced phenotype switching in cells is feasible. Such simulation tools are needed to analyze the functioning of the regulatory circuits that control developmental processes, the mechanisms of bacterial infection, and intercellular signaling.

Harley McAdams
Infernus
724 Esplanada Way
Stanford CA 94305
telephone: (650) 858-1864
fax: (650) 858-1886
email: mcadams@cmgm.stanford.edu

Presentation format: Platform

Gene function: A model for the representation of gene function in a database, FlyBase

Genes are expressed in temporaly and spatially characteristic patterns. Their products are (often) located in specific cellular compartments. These products may be components of complexes. They possess one or more biochemical or physiological functions. These are attributes of genes which are of great interest to all biologists. We need a way to describe these attributes in a rigorous way that will enable biologists to annotate genomes and to explore the universe of genomes. In the ideal world sequence (nucleic acid, protein) and genomic databases would all agree on how this should be done.

For the purposes of this discussion I will discuss the following attributes of genes

- where is a gene expressed ? the spatial problem, described in terms of an organism's anatomy.
- what is the (sub)-cellular localisation of a gene product ?, described in in terms of subcellular anatomy.
- when is a gene expressed ? the temporal problem, described in terms of an organism's ontogeny.
- of what larger structure is a gene product a component ?, described in terms of part-lists of the larger structures.
- what is the function of a gene product ?, described in terms of functional class, metabolic reactions and/or biochemical, cellular, physiological process.

I will illustrate one potential solution to the description of gene function designed for implementation in FlyBase, a comprehensive database of genetic and molecular data concerning Drosophila.

Michael Ashburner
European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 1SD
England
telephone: +44-1223-494648
fax: +44-1223-494468
email: ashburner@ebi.ac.uk

Presentation format: Platform

A report on the Drosophila EST project

The Drosophila EST project is sequencing cDNA clones from embryonic, head, and ovarian libraries that have been enriched for full-length mRNAs. Complete sequences for a smaller number of clones from a membrane enriched library have already been completed. The project was initiated in late 1996 and as of August over 14,000 ESTs were submitted to dbEST. The goal is to generate a minimum of 40,000 high-quality 5' ESTs. We estimate that this will identify 70% of the 12,000 genes in Drosophila. These ESTs are being used to augment the annotation of the genomic sequence the Drosophila Genome Center is generating. The amount of genomic sequence presently stands at ~8.5MB from a genome of approximately 140MB.

In order to derive the intended benefits of this data several key issues need to be addressed. These include computational normalization, assembly of consensus sequences, identification, and correspondence to genomic sequence. We are using a combination of techniques to identify genes. BLASTN and TBLASTX (Altshul et al. 1990) searches are run on both the EST and genomic sequence data against a number of different datasets. The results are interpreted using a program called BOP written here at the center. Genie (Kulp, Reese, and Haussler) is used for gene prediction. The correlation of these analyses and a summary of the results will be presented.

Suzanna Lewis
University of California Berkeley
539 Life Sciences Addition
Berkeley, CA 94720
telephone: 510 643-0269
fax: 510 643-9947
email: suzi@fruitfly.berkeley.edu
url: fruitfly.berkeley.edu

Presentation format: Platform

Functional genomics of the flightless region of *Drosophila*: Implications for genome projects and human diseases

R. Maleszka, H.G. de Couet, G.C. Webb, George L.Gabor Miklos

presenter: George L. Gabor Miklos

The flightless region of *D. melanogaster* has been characterized by genomic and cDNA sequencing, reverse transcription-PCR, deletion analysis, transgenic rescues, and phenotypic dissections. It contains 12 genes, 5 of which have close human relatives, with the remaining 7 having human ESTs. Some are associated with mutant phenotypes; the human flightless homolog (in the Smith-Magenis deletion), the human SOX9 gene (campomelic dysplasia and sex determination), the human dodo homolog (cell cycle abnormalities), the human proline oxidase gene (type 1 hyperprolinaemia), and the mouse diff6 family (defects in cytokinesis). Most of the 12 multidomain proteins are absent from bacteria and half are absent from Saccharomyces cerevisiae. These multilevel data have significant implications for the transferability of functional genomics from model organisms to human beings.

George L Gabor Miklos
The Neurosciences Institute
10640 John Jay Hopkins Drive
SAN DIEGO, CALIFORNIA 92121
telephone: 619 626 2000
fax: 619 626 2079
email: miklos@nsi.edu

Presentation format: Platform

Structure/function analysis of vertebrate Na+/H+ exchanger genes and their role in control of cell volume

We are using a fish model in order to understand the control of red cell volume by the activation and regulation of vertebrate Na+/H+ exchangers (NHE's). Given the conservation of cell volume regulatory behaviour throughout the vertebrates, lessons learned from teleost NHE's will have wider relevance to the control of transport systems in general. In the current study, NHE1 homologues have been isolated from erythrocyte cDNA libraries representing three teleost species whose NHE's have been shown to be regulated very differently (1,2). We are in the process of demonstrating functional expression of these DNA sequences by transfection into the antiporter deficient mammalian cell line PS120. Subsequently we aim to dissect antiporter activity by the construction and expression of a number of inter-specific, chimeric genes. Previous work has shown that in the Trout NHE1 homologue ßNHE, the ability to confer sensitivity to cAMP dependant protein kinase (PKC) lies in the transmembrane region of the gene. A chimera composed of the transmembrane region of human NHE1, which is not PKC sensitive, and the cytoplasmic domain of trout ßNHE, was shown to confer PKC-stimulated volume sensitivity to antiporter deficient PS120 cells (3,4). By comparisons with the sequences that we have isolated and with that of ßNHE, it is apparent that ßNHE has two putative PKC consensus sites in close proximity compared to the remaining sequences which have a single such site. We are interested in the relevance

such sequence diversities have to the functional properties of the NHE protein and the significance in general of the presence/absence of particular consensus sites. For instance, the eel sequence contains a PKC site, suggesting hormonal regulation and yet the experimental evidence so far suggests that such regulation does not take place. We anticipate that identifying the changes in protein structure which mediate sensitivity to different stimuli will also provide insights into the mechanisms by which evolution retailors the functio particular function.

(1)Nikinmaa M., Cech J.J., Ryhanen E-L. and Salama A. (1987) Red cell function of carp (Cyptinus carpio) in acute hypoxia. J.Exptl.Biol 47 53-58.

(2)Romero M.G., Guizoran H., Pellisier B., Garcia-Romeu F. and Motais R (1996) The erythrocyte exchangers of eel (Anguilla anguilla) and Rainbow Trout (Onchorhynchus mykiss): a comparative study. J.Exptl.Biol: 199 415-426.

(3)Borgese F., Sardet C., Cappadoro M., Pouyssegur J. and Motais R (1992) Cloning and expression of a cAMP-activatable Na/H exchanger: evidence that the cytoplasmic domain mediates hormonal regulation. Proc.Natl.Acad.Sci.USA: 89 6765-6769.

(4)Borgese F., Malapert M., Fievet B., Pouyssegur J. and Motais R (1994) The cytoplasmic domain of the Na/H exchangers (NHE's) dictates the nature of the hormonal response: Behaviour of a chimeric human NHE1/trout bNHE antiporter. Proc.Natl.Acad.Sci: 91 5 431-5435.

Dr Cheryl Wright
Department of Environmental and Evolutionary Biology
Derby Building
Brownlow st
Liverpool University
PO box 147
Merseyside L69 3BX
UK
telephone: 0151 794 4985
fax: 0151 794 5094
email: cwright@liv.ac.uk

Presentation format: Poster

An automated digital technology for analyzing expression of nearly all genes

Karl W. Hasel1, Brian S. Hilbush1, Elizabeth A. Thomas2, Jayson Durham1, and J. Gregor Sutcliffe1,2
1Digital Gene Technologies, Inc., 11149 North Torrey Pines Road, La Jolla CA, 92037 and 2The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla CA, 92037

We have developed and automated a method, TOGA (Total Gene expression Analysis), that utilizes a combination of nucleotide sequence and a precise fragment length near the 3' ends of mRNA molecules to give each mRNA in an organism a unique identity, regardless of whether the mRNA has been discovered previously. The identity feature is used in PCR-based assays performed on tissue extracts of interest to determine the presence and relative concentration of nearly every mRNA in the extracts. Using automated DNA sequencing

machines, the data are automatically compiled in a digital form that makes tissue comparisons facile and enables data merging and mining with other information accumulating in a wide range of genome databases and other research resources. A Netscape browser-based graphical user interface exploits these features of TOGATM to allow researchers to quickly identify mRNA expression patterns of interest in experimentally treated, diseased, and control samples, to select candidate mRNAs implicated in disease and drug action paradigms, and to instantly determine whether the particular mRNAs correspond to previously characterized species or are novel.

Karl W. Hasel
Digital Gene Technologies, Inc.
11149 North Torrey Pines Road
La Jolla, CA 92037
telephone: 619-552-1400
fax: 619-552-8625
email: chuck@dgt.com

Presentation format: Platform

Expression mapping of mouse genes

In the last few years there has been a rapid increase in the availability of mammalian gene sequence. In order to begin the process of characterising the transcription of genes and therefore the role of the proteins they encode, we set out to investigate the possibility of mapping the expression patterns of a large number of genes. We have used RT-PCR to assay the expression of genes across a panel of single stranded, oligo dT-primed, cDNAs prepared from 46 individual mouse tissues. Sufficient cDNA was prepared from each RNA sample for the entire study. Primers were designed to the 3' end of genes encoding proteins of widely different functions. To standardise the panel, the amount of cDNA required for each reaction was adjusted so that the amount of PCR product amplified from each tissue was the same when an assay specific for the ribosomal protein S29 was used. Each assay used cDNA equivalent to approximately 10 ng of total RNA and other than the annealing temperature (55 or 60oC) and cycle number (30 to 50), PCR conditions were identical between gene-specific assays. All assays were performed in duplicate to confirm the reproducibility of the results. The profiles observed for different genes ranged from those expressed at an even level across the whole panel, which, on the basis of the tissues studied, appear to have a tissue-independent pattern of expression, to those with a reproducible highly tissue-specific expression profile. To examine the relationship between the RT-PCR profiles and the cellular distribution of mRNA within tissues, we also assayed the expression of a number of the genes by in situ hybridization. Overall there was an excellent correlation between specific areas of sections which exhibited strong hybridization signals, and large amounts of PCR products detected by RT-PCR of the corresponding tissue RNAs. In this study we have determined the expression profiles of over 500 genes and in so doing, demonstrated the method to be robust, sensitive and semi-quantitative. This is the largest single study of the tissue-specificity of gene expression to date, and demonstrates the feasibility of constructing comprehensive expression maps for any mammalian system using RT-PCR. All the expression data is currently being compiled for release on the Jackson Laboratory's Gene Expression Database (GXD). We are now employing this methodology to analyse the expression of patterns of ESTs, aberrant patterns of gene expression associated with disease and for expression profiling of single cells.

Dr Tom Freeman
The Sanger Centre, Wellcome Trust Genome Campus
Hinxton, Cambs., CB10 1SA
UK.

telephone: 0044 1223 494907
fax: 0044 1223 494919
email: tcf@sanger.ac.uk

Presentation format: Platform

SEX Trap: A novel method for the identification of signal sequence-containing genes from cytokine clusters

Miklós Péterfy, Tibor Gyuris, and László Takács

Genes coding for structurally and fuctionally similar proteins are often found in close physical proximity in the genome, forming gene clusters. Examples of cytokine gene clusters include the IL-1 cluster on human chromosome (HSA) 2, the TNF cluster on HSA6, the LIF-Oncostatin M cluster on HSA22, and the interleukin cluster on HSA5.

In order to clone genes of secreted and membrane-bound proteins from selected genomic regions, we combined the principles of signal and exon trapping and developed a new method, SEX trapping. Translation initiated from trapped exons coding for functional signal peptides results in the secretion of a secretory pathway-specific reporter enzyme from COS-7 cells into the cell culture medium. Using test constructs we showed that SEX trapping can identify signal sequence-containing exons from cytokine and non-cytokine genes, and from genomic inserts of widely different lengths. We applied the SEX trap method in the screening of a segment of the interleukin gene cluster on 5q31, which has been partially sequenced at LBNL. Signal sequence-containing exons of the IL-5 and IL-13 genes and a number of potential novel genes have been trapped. Thus, SEX trapping can be a useful tool in the discovery of genes from cytokine and cytokine receptor clusters, as well as in other positional cloning efforts.

Miklos Peterfy
Amgen Center, M/S 8-1-D
Thousand Oaks, CA 91320
telephone: (805)-447-6596
fax: (805)-498-8674
email: mpeterfy@amgen.com

presentation type: Poster

Comparative mapping in the search for the ovine Booroola mutation

Eric. A. Lord, Joanne M. Lumsden, Mike L. Tate, Catherine Y.Y. Liu, Dean J. Burkin and Grant W. Montgomery

AgResearch Molecular Biology Unit, Department of Biochemistry and Centre for Gene Research, University of Otago, PO Box 56, Dunedin, New Zealand

presenter: Eric. A. Lord

The ovine Booroola mutation (FecB) increases ovulation rate and litter size. FecB was mapped to ovine chromosome 6 (OOV6) within a 28 cM region between epidermal growth factor (EGF) and secreted phosphoprotein-1 (SPP1). Since there were no ot her genes mapped to this region of OOV6, we have relied on genetic maps from other species as sources for positional candidates and additional loci to better define the position of FecB on OOV6. This has included genes and ESTs from human chromosome 4, to which OOV6 shares synteny, and candidate genes from the oogenesis pathway in *Drosophila*. Genes and ESTs assigned to HSA4 linkage and physical maps within the critical region were amplified from ovine genomic DNA and a hamster somatic cell hybrid containing OOV6. Primers for six genes were designed from the comparison of mammalian cDNA sequences. Primers pairs from 31 human ESTs were tested and four amplified similar sequences mapping to OOV6. The PCR products were used as RFLP or SSCP probes for linkage mapping in sheep gene mapping flocks and a deer interspecies hybrid mapping panel to confirm their relative positions on OOV6. To improve the efficiency of isolating sequences on OOV6 similar to human ESTs, a further 45 human brain and placental ESTs were screened with a human YAC contig that covers the critical region. The full cDNA clon es of 10 ESTs that mapped to the human YAC contig are being tested as RFLP probes.

Protein sequences of 55 *Drosophila* genes were screened for matches in the EST database and high similarity [P(N) The use of genetic maps from other species has benefited the Booroola mapping programme, firstly, by eliminating known genes and ESTs as potential candidates, and secondly, by adding new loci to the genetic map of OOV6. These additional loci have strengthened the synteny between OOV6 and HSA4. The genetic map in the vicinity of FecB has been better defined to allow us to proceed with physical mapping and transcriptional mapping strategies.

Eric Lord
AgResearch Molecular Biology Unit
Department of Biochemistry and Centre for Gene Research
University of Otago
PO Box 56
Dunedin, New Zealand
telephone: +64-3-4797662
fax: +64-3-4775413
email: lorde@agresearch.cri.nz
url: http://www.agresearch.cri.nz
http://biochem.otago.ac.nz:800/panzora/agresrch.html

Presentation format: Platform

Characterization of a novel anonymous gene in 22q12.2-12.3, a region deleted in sporadic meningiomas

Deletion studies in 170 sporadic meningiomas using a panel of 50 RFLP markers on human chromosome 22 pointed to a large 1 Mbp candidate region on q12.2-12.3 candidate for harboring a new meningioma tumor suppressor gene (Ruttledge et al., 1994). We covered the entire region with overlapping cosmid steps which are currently being sequenced and publicly available at http://www.sanger.ac.uk. We applied a software-based exon-trapping (SBET) procedure, followed by cDNA screening, to all fully sequenced cosmid/BAC clones from the region. This led us to the isolation of a new gene, named V3, ubiquituously expressing a 4700 bp mRNA. Its longest open reading frame is capable of coding for a 756 amino acids protein which does not exhibit any similarity to motifs currently found in protein databases. However, on the amino acid level, it shows 39% identity

to a *C. elegans* putative protein. The most striking feature of this new gene is probably a large genomic size of 400-600 kb. As the genomic sequence covering the entire extend of V3 is not yet fully available, we are in the process of characterizing its genomic organization using the *C. elegans* and the *Fugu rubripes* ortholog genes. Simultaneously, we are testing 170 meningioma cases for rearrangements and point mutations within the gene.

Ref.: Ruttledge, M. H., Xie, Y.-G., Han, F.-Y., Peyrard, M., Collins, P., Nordenskjöld, M. and Dumanski, J. P. (1994). Deletion on chromosome 22 in sporadic meningioma. Genes, Chrom. and Cancer 10: 122-130.

Myriam Peyrard
Department of Molecular Medicine
CMM building L8:00
Karolinska Hospital
Stockholm
S-171 76 Sweden
telephone: +46-8-517 73922
fax: +46-8-517 73909
email: Myriam.Peyrard@cmm.ki.se

Presentation format: Platform

Quantitative analysis of differentially expressed cDNA between normal and deficient thymus lead to the identification of novel functional genes

Catherine Nguyen, Philippe Naquet and Bertrand Jordan

The CD3-z chain of TCR-DC3 complex plays a pivotal role in the activation of T cell responses and in the selection of the T cell repertoire. In z-knockout mice, the T cells have a profound reduction in the surface levels of TCR-CD3 complexes and these animals have poorly developed thymuses. In thymus, the cortex and the medulla are made up of different types of stromal cells that belong to epithelial or hematopoietic lineages. Thymocytes are found in tight interaction with these cells at various stages of differentiation. This complex pattern of migration is precisely regulated in time and space; it likely that part of this process is controlled by specialised cells interaction molecules expressed by stromal cells.

In order to find new genes differentially expressed between normal and z-knock out thymus, we set up two hybridizations on a same set of 3,072 genes with complex probe made from total RNA of z-knock out or wild type thymus. After quantitative measurement of the amount of hybridized probe on each colony, the intensity ratio z-knock out /wild type are calculated. 171 cDNA clones were selected showing either an increased or reduced representation in z-knock out. Additional hybridizations made with complex probes made from RNA of different cells types (such as macrophage, thymocyte, epithelial cell line under different stimulation) or tissues (lymph node, spleen...) allow to refine the selection.

Among clones presenting the most significant profiles, 10 are new or related to mouse EST. Analysed of tissue expression patterns by in situ hybridization shows a selective transcription in epithelial cells types. The sequence analysis reveals that some of these cDNA encode molecules likely involved in the dynamic organisation of thymus.

Catherine Nguyen
"TAGC" CIML INSERM-CNRS
CIML case 906
parc scientifique de luminy
13288
marseille cedex9
France
telephone: 33 4 91 26 94 82
fax: 33 4 91 26 94 30
email: nguyen@ciml;univ-mrs.fr

Presentation format: Platform

Growth failure in idiopathic short stature and Turner syndrome is caused by haploinsufficiency of a pseudoautosomal homeobox gene

Ercole Rao*, Birgit Weiss*, Beate Niesler*, Maki Fukami*,2, Tsutomu Ogata2 and Gudrun A. Rappold*
*Institute of Human Genetics, Heidelberg University, Im Neuenheimer Feld 328, 69120 Heidelberg, Germany. 2 Department of Paediatrics, Keio University, 35 Shinanomachi Shinjuku, Tokyo 160, Japan

A major locus involved in linear growth has been implicated within the pseudoautosomal region (PAR1) of the human sex chromosomes. Cytogenetic studies have provided further evidence that terminal deletions of the short arms of either the X or the Y chromosome (0-700 kb) consistently lead to short stature (SS). We have constructed a cosmid contig across this region. The resulting map was used to position four breakpoints thereby reducing the critical interval for short stature to a 170 kb DNA segment. Using cDNA selection, exon amplification, and CpG island cloning, three novel genes were identified. To search for transcription units within the smallest 170 kb critical region, cDNA selection and exon amplification on six cosmids was carried out. cDNA selection on 25 different cDNA libraries proved to be unsuccessful, suggesting that genes in this interval are expressed at very low abundancy, below the sensitivity level imposed by the method. Exon amplification using the pSPL3-vector allowed the isolation of a homeobox containig exon. The low efficiency of trapped exons was due to the high number of redundant clones generated by cryptic splice sides within the vector. A new approach that we called EASE (exon amplification and selective enrichment) increased the efficiency of positive clones 20fold. Using a pSPL3 derivative vector (pSPL3b) yealded in a further enrichment of positive clones (177fold) and resulted in the isolation of three exons of a novel homeobox-containing gene, SHOX (short stature homeobox-containing gene), within the SS interval. This gene is alternatively spliced encoding proteins with different expression pattern. Mutation analysis and DNA sequencing were used to demonstrate that short stature can be caused by mutations in SHOX.

Ercole Rao
Institut für Humangenetik, INF 328, Heidelberg
Heidelberg, Germany 69120
telephone: 0049 6221 565067
fax: 0049 6221 565332
email: Ercole_Rao@krzmail.krz.uni-heidelberg.de

Presentation format: Platform

A bacterial artificial chromosome expression system for scanning large DNA segments for functional elements: Application in neuroblastoma

Transfer of part or whole chromosomes into tumor cells have previously been successfully used to associate certain chromosomal regions with tumour suppressive activity. However, these studies have fallen short of defining a manageable DNA segment responsible for the activity. We have adopted a bacterial artificial chromosome (BAC) expression vector system, as a tool to scan candidate genomic intervals for their ability to restore non-malignant phenotype in neuroblastoma cells grown in culture. To facilitate selection of stable cell clones, a green fluorescence protein (EGFP) marker, an antibiotic selection gene and a eukaryotic origin of replication were included in the BAC vector. Transfection of large insert BAC clones was performed either by an adenovirus/polyethenimine mediated approach or lipofection. We generally observed 8 % transfection efficiency for 90 kb BAC clones, irrespective of the transfection method used. Stable cell clones were generated after 2 weeks. More than 80 % of the NGP and 100 % of the SK-N-AS stably transfected cell lines continued to express the EGFP marker protein after 12 weeks of passage in culture. Importantly, the BACs were maintained episomally, thus preventing undesired integration into the host cell genome. Evidence demonstrating the applicability of the system in neuroblastoma cells will be presented. The advantage of the system is that large DNAs can be assayed for function, especially where morphological changes are expected. Moreover, generation of a genomic library utilizing this vector would provide an invaluable tool to scan the entire genome for functional elements.

Patrick Onyango
Research Institute of Molecular Pathology (IMP),
Dr. Bohr-gasse 7, A-1030 Vienna, Austria
telephone: 0043 797 30 423
fax: 0043 1 798 71 53
email: Onyango@nt.imp.univie.ac.at

Presentation format: Platform

Database-assisted large scale polymorphism finding and exploitation

Many 'in silico' analyses can now be performed in large human sequence databases. Additionally, these databases can be helpful in the designing of optimised experimental strategies. We are using this latter approach towards the large scale testing of transcribed sequence alleles for associations with complex human disease phenotypes. Allele association studies are usually performed on a 'one gene at a time' basis. Since this is essentially a candidate gene ('guesswork') strategy, then a 'one at a time' effort will not provide an effective way forward. To scale up the system will require three advances, i) multiple intragenic polymorphisms must be identified, ii) facile screening systems must be developed, and iii) appropriate clinical resources must be collected. We are tackling each of these problems.

Intragenic Single Nucleotide Polymorphisms (SNPs) are being gathered by database screening plus automated PCR + sequencing protocols. The practical target is 10,000 SNPs within 18 months. This fundamental aspect of the research is completely dependent upon, and its structure determined by, the wealth of current database information. Details about this aspect of the research shall be presented, and examples given from its application to nuclear genes of oxidative phosphorylation. A one-step, microtitre plate formatted, fluorescence based assay for SNP screening is under development. This will provide automated genotype read-outs for 96 (and perhaps

384) samples at a time. Finally, various clinical collaborations, particularly exploiting a vast and superbly documented Swedish twin registry, will provide appropriate patient and control materials for testing.

Once this system is fully in place, association studies will be possible on a meaningful scale. Furthermore, linkage studies may be performed at far higher speeds and lower costs than with conventional microsatellite analysis. The entertaining problem then will be how to most effectively examine and interpret the large volumes of genotype data produced!

Anthony J. Brookes
Uppsala University
Department of Medical Genetics
Biomedical Centre
Box 589
S-751 23 Uppsala
Sweden
telephone: +46 (18) 471 4151
fax: +46 (18) 526 849
email: tony@medgen.uu.se
url: http://www.medgen.uu.se/

Presentation format: Platform

mRNA architecture and cascade of gene expression

Stable closed regions and flexible open regions were found on mRNAs by analysis of predicted optimal structures and by the ability of closed regions to cause pausing of DNA polymerase during a RT-PCR process. To further substantiate the description of RNA architectural elements, phylogenetic analyses were performed. The CD4 mRNAs of human and several subhuman primate species were analyzed for their predicted secondary structures. The number, location, energy content, energy density of architectural regions from each of the mRNAs were defined. The base pairings and sequence organization shown a significant degree of homology among the CD4 mRNAs of human and subhuman primates. This evolutionary conservation supports that the homologous base-pairings are integral structural components of CD4 mRNA.

The ability of this RNA architecture analysis to predict the effect of antisense oligonucleotide on inhibition of mRNA in cell cultures were examined. The jellyfish green fluorescence protein (GFP) mRNA was used as a target mRNA as it provides an in vivo, real time marker for mRNA degradation as reflected by changes in fluorescence intensity. The architectural regions of GFP mRNA were analyzed and confirmed by RT-PCR assay. Antisense oligonucleotides were designed towards different regions of GFP mRNA and the changes in fluorescence were monitored by flow cytometry. Antisense inhibitors directed towards regions of low energy or open regions were shown to exhibit higher degree of inhibition than those directed towards regions of high energy values. This observation suggests that the efficacy of antisense inhibition can be improved by targeting regions of target mRNA of low energy values.

An application of this observation is to examine the intracellular action of regulatory genes by antisense inhibition and/or by over-expression to define the responsive genes. The responsive genes were then characterized changes in hybridiztion patterns on cDNA arrays and by changes in band intenisty of a RT-PCR process. Some of the responsive genes could futher control the expression of other genes as antisense inhibitors against several responsive genes were shown to alter the expression of other genes. Therefore, the responsive

genes can be classified into group I and group II linked together in a cascade pattern of gene expression.

Wai-Choi Leung, Ph.D.
Tulane University School of Medicine
Dept of Pathology, 1430 Tulane Ave,
New Orleans, LA 70112
telephone: 504-588-5237
fax: 504-587-7389
email: wcleung@tmc.tulane.edu

Presentation format: Platform

Use of cDNA microarrays for analysis of gene expression for thousands of genes simultaneously

D.T.Ross1, M.Eisen2, D. Lashkari3, G. Shuler4, M. Boguski4, J. Hudson5, D. Botstein2, D. Shalon3, P.O. Brown1

Departments of Biochemistry1 and Genetics2, and Howard Hughes Medical Institute1, Stanford University, Stanford CA, Synteni Inc. Fremont CA3, NCBI, Bethesda, MD4, Research Genetics, Huntsville, AL 5.

DNA microarray (chip) technology and two color fluorescent hybridization allows the efficient quantitative determination of relative gene expression on a large scale. We have generated microarrays that contain 10,000 human cDNAs robotically spotted onto treated microscope slides. For comparative hybridization, probe is generated by reverse transcription of mRNA derived from two samples each in the presence of distinguishable fluorescently tagged nucleotides. The probes are combined and hybridized in a small volume (10ul) underneath a coverslip using conventional hybridization chemistry. The arrays are read through use of a custom built confocal laser scanner that generates both a digital reconstructed image, and a quantitative measurement of gene expression in units of relative fluorescent intensity. Examples of the use of the arrays for identification of novel targets of oncogenic proteins, for determination of patterns of gene expression change occurring during induced differentiation,

Douglas T. Ross
Stanford University
Beckman Center B-435
Dept. of Biochemistry
Stanford, CA 94305-5307
telephone: 650-723-6719
fax: 650-725-6044
email: dross@cmgm.stanford.edu

Presentation format: Platform

Development and use of the reverse two-hybrid system to characterize interactions between c-Rel and its inhibitor, IkBa

The yeast two-hybrid system has provided a powerful experimental approach for the identification and characterization of protein:protein interactions. An important feature of the yeast two-hybrid system is the provision for genetic selection techniques that require specific protein:protein interactions. We have developed a modification of the yeast two-hybrid system which enables genetic selection against specific protein:protein interactions. Our reverse two-hybrid system utilizes a yeast strain that contains a mutant cyh2 gene and is therefore resistant to cycloheximide. A wild-type CYH2 gene that is driven by the Gal1 promoter was stably integrated into the genome of this yeast strain. Expression of the wild-type Gal4 protein activates expression of the Gal1 promoter and restores cycloheximide sensitivity. Cycloheximide-sensitive growth can also be restored by coexpression of the wild-type c-Rel and IkBa proteins as Gal4 fusion proteins. Restoration of cycloheximide sensitivity requires assocation between c-Rel and IkBa. Mutant c-Rel proteins can be selected on the basis of their failure to associate with IkBa. The ability to select against specific protein:protein interactions may provide a valuable tool for the functional analysis of proteins.

Mark Hannink
Biochemistry Department, University of Missouri
M121 Medical Science Building
One Hospital Drive
Columbia, Missouri 65212
telephone: (573)-882-7971
fax: (573)-884-4597
email: bcmarkh@muccmail.missouri.edu

Presentation format: Platform

FDD, a high-throughput message display system for the functional interpretation of yeast and mammalian genomes

Genome projects are uncovering a number of novel genes from our genomes as well as those of various model organisms. However, a sizable portion of these genes lack any clues to functions in their structures. It is thus necessary to systematic ally collect biological information other than primary structures for functional interpretation of genome data flooded with such enigmatic genes. Highly informative data would be their expression patterns, disrupted or overexpressed phenotypes, and the mutual relationships.

We have been pursuing the use of so-called message display technology in high-throughput transcript scanning, and established Fluorescent Differential Display system (FDD), that can survey tens of thousands of cDNAs a day. To simultaneously address the three issues described above, we are currently applied FDD to the analysis of yeast gene disruptants/overexpressers for the elucidation of functional relationships between the disrupted/overexpressed gene and those whose transcripts are modulated. To facilitate the generation of expression mutants, we developed a PCR-based promoter replacement strategy to make the expression of any genes under an artificial control. Also, the introduction of an ideal primer set and the multicapillary gel electrophoresis system are planned to accelerate the FDD analysis further.

Another novel intriguing application of FDD is the Allelic Message Display (AMD) for multiplexed imaging of allelic expression status. With AMD among two mouse strains and reciprocal F1 hybrids as well as backcrossed progenies, we have developed a novel screening method for monoallelically expressed transcripts to hunt genes subjected to genomic imprinting, a unique interpretation mode of mammalian genomes. Identification of a novel paternally expressed gene will be presented as a successful example of AMD approach.

Takashi Ito
Human Genome Center, Institute of Medical Science, University of Tokyo
4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan
Tokyo 108
Japan
telephone: 81-3-5449-5623
fax: 81-3-5449-5445
e-mail: tito@ims.u-tokyo.ac.jp

Presentation format: Platform

Comparative analysis and genomic structure of the ataxia telangiectasia gene in human and pufferfish and characterization of some founder mutations

Udar N.S. 1, Morrison A. 2, Telatar M. 3, Cisler A. 2, Amemiya C. 4, Concannon P. 2, Wang Z. 3, Liang T. 3, Chun H. 3, Small K. 1, and Gatti R.A. 3
1Jules Stein Eye Institute, UCLA School of Medicine, Los Angeles, CA 90095. 2Virginia Mason Research Center, Department of Immunology, UW School of Medicine, Seattle, WA 98101. 3Department of Pathology, UCLA School of Medicine, Los Angeles, CA 90095. 4Center for Human Genetics, Boston University School of Medicine, Boston, MA 02118

Ataxia telangiectasia (AT) is an autosomal recessive disease characterized by progressive cerebellar ataxia, immunodeficiency, predisposition to cancer, chromosomal instability and radiosensitivity. Using positional cloning, the gene was localized to a less than 500kb interval on 11q22-23 and subsequently cloned. The gene has PI 3 kinase, rad3 and SH3 domain in addition to a leucine zipper. Recent studies suggest that the ATM protein is involved in signal transduction and forms part of the synaptonemal complex during meiosis. The gene shows strong homologies to TEL1, ESR1/MEC1 genes from yeast. The human ATM gene is contained within ~180kb of genomic DNA. It has 66 exons and the 13kb transcript encodes a protein that is 3056 a.a. and 350kDa. The mouse homologue (Atm) has 84% amino acid identity and 91% similarity to the human counterpart.

We have isolated and sequenced the pufferfish homologue from a lambda and PAC genomic library. Sequence comparison shows a strong conservation in the kinase domain and the 3'end but weaker at the 5'end. It is contained within a 16kb genomic region. Most of the exons and the splice junctions are well conserved.

Our mutation analysis of >280 homozygotes using PTT, SSCP, CSGE, HA and direct sequencing, have identified >150 mutations. PTT detects 70% of ATM mutations. Almost all common mutations were due to founder effect. One public mutation was found in two American families, one of Ashkenazi Jewish background and the other not. Some of the interesting mutations we came across were : 1) Mutations within an intron that create a new splice site, 2) Mutations within an exon that create a new splice site and delete the subsequent exon, 3) Leaky mutations that give more than one mutant RNA, 4) Both in frame and frame shift mutations. We now plan to align the sites of these human mutations in the pufferfish homologue and try to determine new functional domains.

Dr. Nitin S. Udar
Jules Stein Eye Institute

3-544 DSERC
100 Stein Plaza
UCLA School of Medicine
Los Angeles, CA 90095
telephone: 310 794 7420
fax: 310 794 7904
email: NUDAR@Pathology.Medsch.UCLA.Edu

Presentation format: Platform

Evolutionary history of genes from the G6PD region of human Xq28: Insights from the analysis of human; fugu and amphioxus homologues

Z. Sedlacek, E. Steck, J. Coy, and A. Poustka
German Cancer Research Center, Division of Molecular Genome Analysis, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

We isolated and analysed human, fugu and amphioxus homologues of several genes located in the G6PD region of human Xq28 and compared the degree of homology of their nucleotide and amino acid sequences, exon/intron structures and overall gene sizes. Two human, two fugu and one amphioxus rab GDI genes were studied in greatest detail. Their analysis indicated that the occurrence of two forms of rab GDIs preceded the fish-tetrapod divergence, while the amphioxus rab GDI may have evolved from the ancestor of both forms. The gene structures of all rab GDI genes studied were highly conserved. While all were generally composed of 11 exons, the amphioxus gene had one additional intron and one of the fugu genes was missing one intron. The Xq28-located human rab GDI alpha gene was 6.3 kb long and its size was similar to both fugu rab GDI genes. The autosomal human rab GDI b occupied 50 kb of genomic DNA and contained many retroelements in its introns. This size difference reflected the location of the two genes in different isochores of the human genome. The size of the amphioxus rab GDI gene was intermediate to the above genes.

The rab GDIs and several other anchor genes from our model region of Xq28 served us as starting points in the isolation and analysis of large genomic segments in the search for possible traces of ancestral chordate genomic segments or paralogous regions in vertebrate genomes arisen by genome duplications. The data indicate only partial conservation of gene colinearity within the segments studied.

Annemarie Poustka, Barnard Korn, Stephen Wiemann
German Cancer Research Center
Division of Molecular Genome Analysis
Heidelberg 69120
Germany
telephone: +49 6221 42 4702
fax: +49 6221 42 4704
email: s.wiemann@dkfz-heidelberg.de
url: http://www.dkfz-heidelberg.de/abt0840/

X chromosome cDNA library: Generation, evaluation, and application

in transcriptional mapping

B. Korn1,2, S. Wiemann2, H. Roest-Crollius3, H. Lehrach1,3, A. Poustka1,2

1 Resource Center of the German Genome Project 2 Division of Molecular Genome Analysis, German Cancer Research Center, Heidelberg 3 Max-Planck-Institute for Molecular Genetics, Berlin-Dahlem

The technology of cDNA selection has been applied to whole chromosomes through the use of chromosome specific cosmid libraries. We made use of mRNAs from 25 different human tissues as primary, non-cloned cDNA (adrenal gland, adult brain, adult skeletal muscle, bone marrow, fetal brain, fetal kidney, fetal liver, heart muscle, kidney, liver, lung, lymph node, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, small intestine, spinal cord, spleen, stomach, testis, thymus and uterus), in order to have the most complex starting material. We subjected that cDNA to fragmentation and 3' end specific amplification to have access to all genes, not biased by their transcript sizes. This cDNA source was hybridised to the LLNLXU cosmid library, covering the X chromosome (3.7x coverage), in liquid. 23.000 positively selected cDNA clones were picked into microtiter plates and all mapped genes and EST from the human chromosome X were added to this clone collection. High density filters were generated and used for further analysis. Filters are made available through the Ressource Center of the German Genome Project. Randomly chosen clones were sequenced and mapped, either in silico or onto high density filters containing arrayed X chromosome specific YACs . In silico mapping was supported by the fact that we designed the cDNA selection to specifically clone the 3' end of transcripts and, therefore, have immediate access to the EST databases by BlastN alignment. Results of the sequence analysis of 94 randomly chosen clones are given below.

Table 1

no database hits	39,36%
ESTs of unknown location	30,85%
Genes/ESTs mapped to human X	18,09%
repetitive sequences	5,32%
ESTs/genes located on non-X chromosomes	4,26%
background (e. g. E. coli,...)	2,13%

By this approach we have been able to quality control the established X chromosome cDNA library and to identify and map new genes at the X chromosome.

Annemarie Poustka, Barnard Korn, Stephen Wiemann
German Cancer Research Center
Division of Molecular Genome Analysis
Heidelberg 69120
Germany
telephone: +49 6221 42 4702
fax: +49 6221 42 4704
email: s.wiemann@dkfz-heidelberg.de
url: http://www.dkfz-heidelberg.de/abt0840/

Generation and sequencing of full length cDNAs in the course of the

German Genome Project

Stefan Wiemann(1), Bernhard Korn(1,2), Annemarie Poustka(1,2)
Division of Molecular Genome Analysis(1)
Resource Center of the German Genome Project(2), German Cancer Research Center, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany

In the frame of the German Genome Project, we generate cDNA libraries enriched for 'full length' cDNAs from human tissues with the aim to obtain complete cDNAs, from the 5' cap structure to the poly A tail of as many as possible human genes. Today libraries have been made from fetal brain, fetal kidney, testis, skeletal muscle and spinal cord. Starting from oligo-dT primed cDNA, the first strand cDNA is amplified under long range conditions in few cycles. Cloning is done directionally into plasmid vectors. Libraries (30,000 to 120,000 clones) are picked in 384 well plates and spotted on high density filters. We initially characterize new libraries by sequencing randomly picked clones and hybridization of genes of varying size and abundance. The libraries have already been successfully used to screen for full length representations of the human MTM1 gene (3.4 kb), the human homologue of flightless (4.2 kb,) and a number of other transcripts. Currently, the libraries are used to isolate full length representations of the genes located in the chromosomal region Xq27.3 - qter.

Clones with inserts longer than 1.5 kb are pre-selected by agarose gel electrophoresis, for subsequent sequence analysis. Highly abundant genes are hybridized to the filters in order to minimize redundancy in initial EST sequencing. EST sequences are analyzed for the likelihood of the clones to be full length, e.g. by the presence of CpG clusters, in order to obtain a minimal set of full length clones for efficient complete sequence analysis. Clones identified to be full length are sequenced and further analyzed by a national consortium in the frame of the German Genome Project. We will determine 8 Mb of finished sequence comprising 3,000 - 4,000 full length cDNA sequences in the next three years. The sequences are analyzed for possible function in silico to facilitate subsequent functional analysis. All clones and data generated during the project are made publicly available via the Resource Centre of the German Genome Project (RZPD).

Annemarie Poustka, Barnard Korn, Stephen Wiemann
German Cancer Research Center
Division of Molecular Genome Analysis
Heidelberg 69120
Germany
telephone: +49 6221 42 4702
fax: +49 6221 42 4704
email: s.wiemann@dkfz-heidelberg.de
url: http://www.dkfz-heidelberg.de/abt0840/

Presentation format: Platform

Structural and compositional features of untranslated regions of eukaryotic mRNAs

The important role of 5' and 3' untranslated regions of eukaryotic mRNAs in gene regulation and expression is now widely acknowledged. In order to study the general structural and compositional features of these sequences we developed UTRdb, a specialized database of 5' and 3'-UTR sequences from seven different taxonomic groups of eukaryotic mRNAs cleaned of redundancy. UTRdb (release 4.0) contains about 60,000 entries and

18,500,000 nucleotides. The analysis of the UTR sequences contained in this database showed that 5'-UTR sequences, on average 200 nucleotides long, are 3 to 1.5 times shorter than corresponding 3'-UTR sequences in the various taxonomic groups considered here. As far as the compositional properties are concerned, on average 5'-UTR sequences resulted in all cases GC richer than 3'-UTR sequences and significant correlations were found between the GC content of 5' and 3'-UTR sequences and the GC content of the third silent codon positions of the corresponding protein coding genes. Dinucleotide analysis showed a differential depletion of CpG in vertebrate 5' and 3'-UTR, with 5'-UTR sequences CpG richer. A generalized depletion of TpA in both 5' and 3'-UTR was observed in all eukariotic sequence collections. Furthermore, by using suitable algorithms we searched UTR sequences for primary and/or secondary structure motifs possibly endowed of some biological role in gene regulation and expression.

Graziano Pesole
Dept. of Biology D.B.A.F.
University of Basilicata, Italy
via Anzio 10
Potenza 70126
Italy
telephone: +39-971-474431
fax: +39-971-474439
email: graziano@area.ba.cnr.it

Presentation format: Platform

Expression pattern map of the *C. elegans* genome

Aiming at understanding of the gene expression networks in development of *C. elegans*, we are constructing an expression pattern map of the 100Mb genome through identifying and characterizing cDNA clones of all the genes whose total number is estimated to be 15,000.

Thus far, some 40,000 clones have been subjected to tag-sequencing from both ends, which were classified into about 7,300 unique cDNA species (almost half of the total genes) by comparing the 3'-tags. Most of them were mapped on the genome either by in silico mapping or hybridization to the YAC filters. We are systematically analyzing the expression patterns of the classified cDNA species using in situ hybridization on whole mount specimen of embryos, larvae and adults. Rough estimation is that 39% of the cDNA species show specific pattern of expression during embryogenesis, out of which 1/3 shows zygotic expression and 1/3 maternal expression. mRNA of 1/4 of the maternally expressed genes (about 4% of all) disappears in very early stages before gastrulation. Classifying of the expression patterns leads to identification of sets of genes which show the very similar expression patterns. Determination of the regulatory regions of these genes is also in progress. We are constructing the database named NEXTDB (the Nematode EXpression paTtern DataBase) to make the information available on the internet.

Yuji Kohara
Gene Network Lab
National Institute of Genetics
1111, Yata
Mishima 411

Japan
telephone: +81-559-81-6854
fax: +81-559-81-6855
email: ykohara@lab.nig.ac.jp

Presentation format: Platform

Gene expression resource for the laboratory mouse

M. Ringwald, R. Baldock*, J. Bard#, D. Begley, G. Davis, D. Davidson*, J.T. Eppig, K. Frazer, M. Kaufman#, M. Mangan, J. Richardson, L. Trepanier

The Jackson Laboratory, Bar Harbor; *MRC Human Genetics Unit, Edinburgh; #Edinburgh University, Edinburgh.

The process of differential gene expression generates extraordinarily complex spatio-temporal networks of gene and protein interactions. With its new focus on gene function and expression analysis, genome research is beginning to elucidate these networks to understand the molecular basis of human health and disease. The laboratory mouse will serve as a pivotal animal model in these studies. High throughput expression methods will make it possible to analyze in parallel the expression of thousands of genes in different tissues that can be derived from many different mouse strains and mutants. These experiments will provide global insights into expression profiles and molecular pathways, and lead the way to more focused expression studies using Northern and Western blot, RT-PCR, RNA in situ hybridization, and immunohistochemistry assays to determine what transcripts and proteins are produced by specific genes, and where and when these products are expressed at the cellular level.

We are developing a database of gene expression information for the laboratory mouse that can store and integrate these data and make the data freely and widely available in formats appropriate for thorough analysis. Expression patterns are described using a standardized anatomical dictionary. For in situ studies, the textual annotations are complemented with digitized images of original expression data that are indexed via the terms from the dictionary. This database system will be combined with a 3D atlas of mouse development to enable 3D graphical display and analysis of expression patterns. Integration with the Mouse Genome Database and comprehensive interconnections with other relevant databases will place the gene expression data into the larger biological and analytical context.

Expression data are and will be acquired from the literature by database editors, but primarily data will come via electronic submissions directly from research laboratories. The Gene Expression Annotator, an electronic submission system for expression data that we have developed, is currently being tested by a number of laboratories in North America and Europe with the aim of developing a user friendly system for the community at large. The Gene Expression Index, a searchable index into the expression literature for mouse development, being updated daily, is already accessible to the general public at http://www.informatics.jax.org/gxd.html. Additional data sets will be made available in the near future. The current status of the database and its future applications will be discussed.

Martin Ringwald
The Jackson Laboratory
600 Main Street
Bar Harbor, Maine 04609

telephone: (207) 288-6436
fax: (207) 288-6132
email: ringwald@informatics.jax.org

Analysis of differentially expressed genes using a solid-phase RDA approach

Solid-phase methods based representational differential analysis (RDA) have been designed enabling differential gene expression analysis in samples with scarce amounts of mRNA originating from skin and colon tissue. A microdissection procedure has been developed in parallel for analysis of small cell cluster in tissue sections using a laser-assisted capture microscope. This procedure of selection of specific cell populations has been combined with the solid phase RDA principle employing the streptavidin biotin system to capture nucleic acids onto microbeads for further use in vitro amplification systems. The immobilisation of nucleic acids to a solid phase has significantly simplified the purification process and minimised sample loss that may also facilitate future automation.

Joakim Lundeberg
Royal Institute of Technology
Department of Biochemistry
KTH-Royal Institute of Technology
Stockholm S-100 44
SWEDEN
telephone: 46 8 790 87 58
fax: 46 8 24 54 52
email: joakim.lundeberg@biochem.kth.se

Presentation format: Platform

Attaching functional annotation to predicted genes in genomic sequences

As the Human Genome Project enters the large-scale sequencing phase, computational gene identification methods are becoming essential for the automatic analysis and annotation of large uncharacterized genomic sequences. Substantial progress has been made in the recent years in the field of computational gene identification, and when the location of the genes in the genomic sequences is approximately known, computer programs exist that are able to predict the exon/intron boundaries with high accuracy. However, currently available programs are still unable to succesfully cope with anonymous sequences a few megabases long containing an unknown number of genes---the sequences typically produced in the large Genome Centers. Moreover finding the genes and deciphering gene structure is only the first step towards the automatic annotation of genomic sequences; attaching relevant functional information to the predicted genes is also essential. Here, we will discuss recent developments in the GeneID program to address both these problems: predicting genes in very long anonymous genomic sequences, and automatically attaching functional annotation to the predicted genes. In particular, we will describe the methodology used to assign functional descriptions to the predicted genes based

on the functional annotation of similar amino acid sequences in the public databases. By means of a process which we term "reverse querying of a database", the first order boolean formula built on the annotation of a protein sequence database is found, that best describes the set of amino acid sequences showing similarity to the amino acid sequence encoded by a predicted gene. Such a formula is assumed to be the best description for the function of the gene. A measure of quality is computed for the descriptions obtained, and thus, the ability to assign a good functional description to a predicted gene may reinforce the confidence in the reliability of the prediction. Functional annotation is also attempted for connected regions of similarity to amino acid sequences along the DNA sequence---which may not be assembled into genes. In cases of low or controversial similarity, the quality of the assigned functional prediction can be used to independently asses the biological significance of the amino acid matches.

Roderic Guigo
Informatica Medica
Institut Municipal d'Investigacio Medica (IMIM)
C/ Dr. Aiguader 80
Barcelona 08003
Spain
telephone: +34 3 221 1009
fax: + 34 3 221 3237
email: rguigo@imim.es
url: http://www1.imim.es/~rguigo/Welcome.html

Presentation format: Platform

A linkage map of zebrafish transcribed sequences and the evolution of the vertebrate genome

To investigate mechanisms of vertebrate genome evolution, we localized 135 transcribed sequences on the zebrafish linkage map and compared results to mammalian gene maps. Analysis revealed large chromosome segments conserved among species. Up to four copies of paralogous chromosome segments exist in zebrafish, and they generally correspond to orthologous chromosome segments in mammals. These results suggest that two polyploidization events occurred in vertebrate evolution prior to the divergence of fish and mammal lineages. An additional round of chromosome duplications may have occurred in the zebrafish lineage. Comparative genomics suggests the content of chromosomes in the pre polyploidization common ancestor of zebrafish and mammals. This zebrafish map will facilitate molecular identification of mutated zebrafish genes, which can suggest functions for human genes known only by sequence.

John H. Postlethwait
Institute of Neuroscience
University of Oregon
Eugene, OR 97403
telephone: 541-346-4538
fax: 541-346-4538
email: jpostle@oregon.uoregon.edu

Presentation format: Platform

Structural and compositional features of untranslated regions of eukaryotic mRNAs

The important role of 5' and 3' untranslated regions of eukaryotic mRNAs in gene regulation and expression is now widely acknowledged. In order to study the general structural and compositional features of these sequences we developed UTRdb, a specialized database of 5' and 3'-UTR sequences from seven different taxonomic groups of eukaryotic mRNAs cleaned of redundancy. UTRdb (release 4.0) contains about 60,000 entries and 18,500,000 nucleotides. The analysis of the UTR sequences contained in this database showed that 5'-UTR sequences, on average 200 nucleotides long, are 3 to 1.5 times shorter than corresponding 3'-UTR sequences in the various taxonomic groups considered here. As far as the compositional properties are concerned, on average 5'-UTR sequences resulted in all cases GC richer than 3'-UTR sequences and significant correlations were found between the GC content of 5' and 3'-UTR sequences and the GC content of the third silent codon positions of the corresponding protein coding genes. Dinucleotide analysis showed a differential depletion of CpG in vertebrate 5' and 3'-UTR, with 5'-UTR sequences CpG richer. A generalized depletion of TpA in both 5' and 3'-UTR was observed in all eukariotic sequence collections. Furthermore, by using suitable algorithms we searched UTR sequences for primary and/or secondary structure motifs possibly endowed of some biological role in gene regulation and expression.

Graziano Pesole
Dipartimento di Biologia D.B.A.F.
via Anzio 10
POTENZA, 85100
ITALY
telephone: +39-971-474431
fax: +39-971-474439
email: graziano@area.ba.cnr.it

Presentation format: Platform

Phenotypic dissection of complex traits in mice: risk factors, expression profiles, physiological and developmental pathways, and models for human genetic diseases

Many important human genetic diseases are genetically and phenotypically complex and often involve combinations of genes that may interact with each other or with environmental factors to cause disease. While formal, rigorous methods have been developed for the genetic dissection of these traits in both humans and model species, phenotypic dissection may prove to be as important as genetic dissection in the end-game of identifying and characterizing candidate disease susceptibility genes in complex traits. However, the paradigms and strategies for phenotypic dissection of complex traits remain relatively poorly developed. We have been exploring ways to address this problem by focusing on particular diseases, mouse models, physiological pathways, biochemical assays, and gene expression profiles. The folate and homocysteine metabolic pathways have several features that make them ideal for these proof-of-concept studies, including their involvement in cardiovascular disease, neural tube defects, colon cancer, and seizures, the likelihood of finding or making mouse

models, the well-characterized metabolic pathway, the availability for assays of enzyme activities and metabolite assays, and the possibility of using expression profiles to characterize the metabolic anomalies. I will illustrate this paradigm with results from surveys and analyses of homocysteine levels, MTHFR activities, and expression profiles in inbred strains and mutant mice. By combining these methods for phenotypic dissection of pathways and complex traits with traditional genetic analyses such as linkage studies, powerful opportunities are now available to make progress in identifying and characterizing new mouse models for common human birth defects and diseases. This paradigm should apply to many other kinds of genetic diseases and physiological pathways.

Joseph H. Nadeau
Genetics Dept., Case Western Reserve University School of Medicine
10900 Euclid Ave
Cleveland, Ohio 44106
telephone: 216-368-0581
fax: 216-368-3432
e-mail: jhn4@po.cwru.edu

Presentation format: Platform

Gene identification in sequenced DNA: Database searches, gene predictions and verification by RT-PCR, Northern blots, and exon trapping

J.T. Den Dunnen, I. Stec, E. Van De Vosse, D. Jennen and G.J.B. Van Ommen

The way in which candidate disease genes are identified in positional cloning projects is rapidly changing. A few years ago, one had to use methods like cDNA-selection, cDNA- library screening , identification of evolutionary conservation sequences, localization of HTF- islands or exon trapping. Currently, an increasing number of "electronic" possibilities are emerging, including the analysis of the mapped ESTs and searching databases for functional candidates. The most recent addition is the analysis of the large stretches of sequenced DNA that are emerging from the Human Genome Project. We are involved in two positional cloning projects, Wolf-Hirschhorn Syndrome (WHS) and Retinoschisis (RS), for which the entire disease gene candidate region has been, or is being, sequenced; 165 kb on 4p13 and 1.2 Mb on Xp22 respectively.

Computer analysis of the sequenced regions revealed several problems. On one hand, the diversity of databases and gene prediction programs available created very helpful resources, but on the other hand, the lack of specifically designed software made the analysis very time consuming. For example, since repeat masking was not perfect, repeated database searches had to be performed. Furthermore, detailed analyses of the results would be simplified if e.g. retrieval of a 5' EST sequence would simultaneously yield the 3' sequence or, even better, the entire batch from a "UniGene" set. dbEST contained the most valuable data resource. For many genes both human and murine transcripts were present. Some ESTs seem to be derived from priming at intronic A-rich regions, others probably derive from hnRNA (or genomic DNA), with A-rich regions at both ends. ESTs from both DNA strands were also detected; in one case this probably derived from a duplicated genomic sequence. Database searches using translations of the constructed putative gene sequences against the six frame translations of dbEST frequently identified transcripts from diverse organisms with high local similarities, probably representing new protein domains.

The electronic results were verified using RT-PCR, cDNA-library screening, exon trapping and Northern blot

analysis. Both Northern analysis and RT-PCR were facilitated by the expression profile deduced from dbEST. RT-PCR was most powerful to link computer- predicted exons and to verify intron/exons borders. Identification of the gene's 5' end turned out to be the most difficult, especially since RT-PCR analysis seemed to link transcripts from directly flanking genes. For one region, containing a large open reading frame which was clearly evolutionary conserved, transcripts could never be identified, neither using RT- PCR, nor using cDNA-library screening or Northern analysis. Furthermore, database searches revealed no homologies.

Dr. Johan T. den Dunnen
Department of Human Genetics
Leiden University
Wassenaarseweg 72
2333 AL LEIDEN
the Netherlands
telephone: +31-71-5276105
fax: +31-71-5276075
email: ddunnen@ruly46.medfac.leidenuniv.nl
url: http://ruly70.medfac.leidenuniv.nl/

Presentation format: Platform

Development of libraries enriched for full-length cDNAs: A progress report

1Maria de Fatima Bonaldo, 2Kala Mayur and 1Marcelo Bento Soares
1Department of Pediatrics and Physiology and Biophysics, The University of Iowa, 2Department of Psychiatry, Columbia University.

Among other applications, the availability of libraries enriched for full-length cDNAs will facilitate all ongoing efforts aimed at the identification of disease-causing genes. A few methods have been described for construction of full-length cDNA libraries, which take advantage of the CAP structure present at the 5' end of intact mRNAs, to select for "full-length" molecules (1-4). However, as a general rule, these libraries have not been characterized to the extent that it would be required to determine whether most clones are truly full-length. It is noteworthy, however, that at the very least these procedures will be most valuable to generate libraries enriched for 5' ends of mRNAs. There are two potential problems. First of all, since there is no selection for bonafide 3' ends, many of the resulting clones may be truncated at the 3' end. This is so because mRNAs may be primed internally during synthesis of first-strand cDNA. Because of their smaller size, these clones may outcompete their full-length counterparts during ligation and amplification. Second, and most importantly, the differential clonability and growth properties of smaller (full-length) cDNAs versus longer (full-length) cDNAs make it very difficult to isolate long full-length cDNAs. In an effort to address these concerns, we developed an alternative strategy for construction of libraries enriched for full-length cDNAs which is based on the rationale that if cDNA is synthesized from size fractionated mRNA, it can be strictly size selected prior to cloning accordingly, thus yielding sub-libraries greatly enriched for full-length cDNAs. We have documented the feasibility of this approach by generating a number of such sub-libraries from a mixture of size-fractionated mRNA from human brain and placenta. The results indicated that the sub-libraries produced were significantly enriched for full-length cDNAs. However, detailed characterization of these libraries also pointed to some problems which we are currently attempting to solve. A critical review of the advantages and disadvantages of this procedure will be presented.

1. Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki, N., Izawa, M., Muramatsu, M., Hayashizaki, Y. and Schneider, C. (1996). High-efficiency full-length cDNA cloning by biotinylated CAP trapper. Geno mics 37: 327-336.

2. Carninci, P., Westover, A., Nishiyama, Y., Ohsumi, T., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., Schneider, C and Hayashizaki, Y. (1997). High-efficiency selection of full-length cDNA by improved biotinylated Cap trapper. DNA Resea rch 4: 61-66.

3. Kato, S., Sekine, S., Oh, S-W., Kim, N-S., Umezawa, Y., Abe, N., Yokoyama-Kobayashi, M. and Aoki, T. (1994). Construction of a human full-length cDNA bank. Gene 150: 243-250.

4. Edery, I., Chu, L.L., Sonenberg, N. and Pelletier, J. (1995). An efficient strategy to isolate full-length cDNAs based on a mRNA Cap retention procedure (CAPture). Mol. Cell. Biol. 15: 3363-3371.

Marcelo Bento Soares
The University of Iowa
451 Eckstein Medical Research Building
Iowa City, IA 52242
telephone: (319) 335-8250
fax: (319) 335-9565
email: bento-soares@uiowa.edu

Presentation format: Platform

Linking human proteins using two-hybrid technology

The GeneNet Project is a special project that uses CLONTECH's unique yeast two-hybrrid approach to constructing a total human genome protein linkage map database. So far we have reached two important milestones: 1) Constructed a nove and complete human ES T-GAL4 AD fusion library. This library constains 250,000 human EST cDNA inserts in all 3 reading frames originally present in Washington/Merck EST Project libraries. This library covers most human tissues and cell types. This library was constructed by a proprietary technique developed at CLONTECH. 2) Screened this library against more than 20 arbitrarily selected human protein genes. The results confirmed previously known interections among some of these proteins. In addition, we have shown a localized protein interaction network that linkes some well-known Bcl-2 family proteins with some TNF receptor-associated proteins, a few well-studied signal transduction proteins, and a newly identifies tumor suppressor protein. These results have demonstrated the potential value of this project: it is possible to build a total human protein linkage map using this approach. The ultimate linkage map will connect all human protein genes, represented by existing and new ESTs, into a well organized 3-dimentional database , which will have tremendous impact on pharmaceutical applications.

Li Zhu
CLONTECH Laboratories, Inc.
1020 E. Meadow Circle
Palo Alto, CA 94303
telephone: (650)-424-8222x1462
fax: (650)-354-0776
url: liz@clontech.com

Presentation format: Platform

Reverse genetics by chemical mutagenesis in C. elegans

Gert Jansen, Karen L. Thijssen, Esther Hazendonk, Marieke van der Horst and Ronald H.A. Plasterk

The nematode *C. elegans* will be the first animal for which the complete genome sequence will become available. This opens the possibility of new large scale functional studies of e.g. whole gene families, using loss of function and gain of function mutants and expression patterns. Thus far the method of choice for gene inactivation was a two-step approach using the transposon Tc1 (Zwaal et al., 1993, Proc. Natl. Acad. Sci. USA 90, 7431-7435). We have developed a one step method for target-selected gene inactivation in C. elegans using chemical mutagenesis (Jansen et al., 1997, Nature Genet. 17, 119-121). A permanent frozen mutant collection has been established, consisting of over 7,000 cultures, each representing approximately 150 genomes. We use PCR to selectively visualize deletions in genes of interest: primers are selected more than 3 kbp apart, so that the deletion fragment will have a selective advantage in the amplification reaction. The method is sufficiently sensitive to permit detection of a single mutant among more than 15,000 wild types. The approach has successfully been applied in our study of the function of all heterotrimeric G-protein genes in C. elegans. We will discuss our plans to scale up the method for systematic inactivation of all 17,000 C. elegans genes.

Gert Jansen
Division of Molecular Biology (H8)
The Netherlands Cancer Institute
Plesmanlaan 121
1066 CX Amsterdam
The Netherlands
telephone: #31-20-5122090
fax: #31-20-5122086
email: GJANS@NKI.NL
url: http://www.nki.nl/nkidep/h8/people/gert.html

Presentation format: Platform

The mouse Sry interactive proteins are differentially expressed in adult and fetal tissues

J.Q. Zhang, P. Coward, M.W. Xian and Y-FC. Lau
Division of Cell and Developmental Genetics, Dept. of Medicine, University of California, San Francisco, California

The mouse testis determining gene, Sry, on the Y chromosome encodes a protein with a DNA-binding (HMG box) domain at its amino end and a glutamine-rich domain at its carboxyl end. The HMG box is conserved in the Sry of other mammals. The glutamine-rich domain is encoded mostly by CAG repeats and its function is unknown. We hypothesize that the glutamine- rich domain was generated by an in-frame insertion of a repetitive sequence in the mouse ancestral Sry. It had gained a protein-protein binding function, similar to situations of the CAG expansion in the mutated genes of several neurodegenerative diseases. However, in the case of the mouse,

an evolutionary adaptation results in a stable retention of the CAG repeats in Sry. Using the glutamine- rich domain as probes in farwestern blotting studies, we detected 3 specific bands at 94, 32 and 28 kDa only in testis extract and a 90 kDa in the brain extract from adult tissues. The 94, 32 and 28 kDa testicular proteins have been designated as Sry interactive protein 1 (Sip-1), 2 (Sip-2) and 3 (Sip-3) respectively. The Sips were detected in somatic cells of testicular origin and their expression was associated with spermatogenic activities. Additional studies using subcellular fractionation techniques demonstrated that both Sip-2 and -3 were predominantly present in the nuclei while Sip-1 was present in both cytoplasmic and nuclear fractions. In situ blotting and farwestern blotting of adult testis section demonstrated that indeed the Sips were preferentially localized in the interstitial and peripheral regions of the seminiferous tubules. Sips were expressed in tissues of embryos as early as 8.5 days post coitus (dpc) and in fetal gonads of both sexes at 11.5 dpc, during the time of sex determination. However, their expression patterns varied both quantitatively and qualitatively at different developmental stages. Although the exact nature of these Sry interactive proteins has yet to be defined, their detection supports our hypothesis that the mouse Sry glutamine-rich domain contributes to the biological function(s) of Sry through a protein-protein interactive role(s).

Chris Lau
Division of Cell and Developmental Genetics
Department of Medicine, VAMC-111C5
University of California, San Francisco
4150 Clement Street
San Francisco, California 94121
telephone: 415-476-8839
fax: 415-502-1613
email: clau@itsa.ucsf.edu

Presentation format: Platform

Gene discovery in AT-rich regions of human chromosome 21 and correlation of base composition with compaction in *Fugu rubripes* homologous genes

1. The Giemsa dark band, 21q21, contains 50% of human 21q DNA (20 Mb), but by mapping of characterized genes, by cDNA selection, and by mapping of dbEST entries, it contains only 10-15% of the genes. These data do not unequivocably rule out the possibility that 21q21 harbors genes with restricted time and/or place of expression, but gene discovery is hampered by two additional features of the region: 1) 21q21 is underrepresented in many libaries and clone contigs, and ii) it is AT-rich and therefore less reliable in analysis by exon prediction programs. We are using exon trapping from cosmids specifically derived from 21q21, and sequence analysis of *Fugu rubripes* cosmids containing 21q21 homologous sequences to aid in gene discovery. Results so far include:

i) Microdissection of 21q21 has provided a number of novel unique sequences; 70 of these have been used to screen the LLNL chromosome 21 specific cosmid library. A nonredundant subset of positive cosmids was then used in exon trapping experiments. Results indicate an approximately 8-fold lower density of putative exons and shows that these are significantly smaller and much less GC-rich than those within 21q22. Sequencing in the parent cosmids indicates that the exon trapped products are likely bona fide exons and novel.

ii) The APP gene from *Fugu rubripes* has been isolated and completely sequenced. Several exon prediction

programs applied to the *Fugu* sequence show a slightly higher true positive rate and a significantly lower false positive rate than is obtained with the same programs applied to the genomic sequence of the human APP gene. Were the human APP protein now known, use of the *Fugu* sequence would reduce time and effort required to verify gene predictions. Conserved synteny also aided in gene discovery: <3 kb downstream of APP in the same *Fugu* cosmid, the GABPA transcription factor was found, a gene in humans known only to lie within 800 kb of APP.

2. Determination of the genomic structure of the *Fugu* APP gene revealed unusually high compaction: the *Fugu* introns are on average 50-fold smaller than the human homologues. APP is also the most AT-rich human gene so far analyzed in *Fugu*. Characterization of additional genes whose isochore location is known in human indicates that base composition can predict the relative extent of compaction of the homologue. This supports the value of *Fugu* analysis, particularly in AT-rich regions.

Katheleen Gardiner
Eleanor Roosevelt Institute
11899 Gaylord Street
Denver, Colorado 80206
303-333-4515
303-333-8423
gardiner@eri.uchsc.edu

Marie-Laure Yaspo
Max Planck Institut fur Molekulare Genetick
Ihnestrasse 73
Berlin D-14195
GERMANY
49-30-8413-1356
49-30-8413-1380
yaspo@mpimg-berlin-dahlem.mpg.de

Presentation format: Platform

RNA editase genes from *Fugu rubripes*

The known mammalian A-to-I RNA editase genes include DRADA and RED1, involved in the editing of glutamate and serotonin receptors, and RED2, with as yet undefined substrates. The genomic structure of the human DRADA gene is known and we have determined the structure of the human RED1 gene. While these proteins share some substrates, their protein structure and intron-exon organization are significantly different. DRADA is composed of 15 exons; RED1, only 10. DRADA contains three RNA binding domains, each split by an intron at a conserved site and contained within exons 2+3, 4+5 and 6+7. RED1 contains only two RNA binding domains and these are entirely contained within an unusually large (>900 nucleotide) exon 2.

A cDNA for human RED1 was used to screen a *Fugu rubripes* cosmid library (constructed by Greg Elgar; archived at the German Human Genome Resource Center), identifying 4 non- overlapping cosmids. Sequence analysis revealed a family of editases. One cosmid contains the homologue of DRADS, identified by the similarity of number and organization of RNA binding doamins, and intron-exon structure. A second cosmid contains the apparent different genes (named RED1a and RED1b), each showing high homology to human RED1. The exon-intron boundaries of RED1a and RED1b are conserved with human, but intron sizes vary, both

from human and between each other. Overall the protein similarity between RED1a and RED1b is 83%; in the two RNA binding domains the similarity rises to >95%.

Together, these data imply that *Fugu* contains at least 4 distinct A-to-I RNA editase genes, suggesting that additional editases remain to be identified in mammals. These *Fugu* genes also provide the material for further characterization of some unusual features observed in the human RED1 3' UTR.

Dobrimir Slavov
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, Colorado 80206
telephone: 303-333-4515
fax: 303-333-8423
email: gardiner@eri.uchsc.edu

Presentation format: Platform

Chromosome 3q breakpoints in leukemia: Complexities in alterntive processing and intergenic splicing

Rearrangements of chromosome 3 in Leukemia include the so-called "3q syndrome" involving t(3;3)(q21;q26) and inv(3) (q21;q25), observed in 4%-6% of acute myelogenous leukemia (AML). Both the 3q21 and the 3q26 breakpoint regions have been studied extensively. In 3q21, we have used gene identification and breakpoint mapping to reveal several unusual characteristics: i) the region is very gene rich, with results from cDNA selection, exon trapping and genomic sequence analysis suggesting one gene per 10 kb over an 80 kb segment; ii) breakpoints are clustered within a 30 kb segment but are dispersed among genes, occurring both 5' and 3' to a number of different genes; and iii) breakpoints both 5' and 3' can activate expression of some of these genes (GR6 and 2C12). In contrast, in 3q26, others have shown that breakpoints are dispersed over several hundreds of kb, both 5' and 3' to the Zn finger transcription factor, EVI1, and >170 kb upstream of EVI1, 5' to the adjacent MDS1 gene.

There are, however, similarities at the expression level between the 3q21 and 3q26 breakpoint regions. These include: i) restricted expression patterns: e.g. in 3q21, expression of the GR6 gene has been observed only in early fetal development; EVI1 exhibits fetal and tissue specificity; ii) complex alternative splicing: e.g. both GR6 and EVI1 genes display several alternative transcripts including those produced by use of splice sites within exons and read through into introns; and iii) intergenic splicing: in some normal tissues, intergenic splicing between the MDS1 and the EVI1 genes in 3q26 is observed. In AML with t(3;3), intergenic splicing between the GR6 and RPBHI genes in 3q21 and EVI1 in 3q26 is observed. The 3q21 breakpoints are up to 30 kb downstream of the 3q21 genes. These unusual features and their implications in normal expression patterns and leukemia will be discussed.

A. Rynditch, Y. Pekarsky, K. Gardiner
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, Colorado 80206
telephone: 303-333-4515
fax: 303-333-8423

email: gardiner@eri.uchsc.edu

Presentation format: Platform

Functional studies of coding and non-coding sequences of zebrafish genome

Studies of relatively simple organisms have yielded much of our current understanding of the molecular mechanisms underlying proliferation, commitment, differentiation, and pattern formation during animal development. Zebrafish are rapidly becoming a popular model organism for genetic studies of these processes. A female zebrafish typically produces up to several hundred transparent embryos that rapidly develop outside the mother. These features make it possible to perform a systematic analysis of genome function, including both coding and noncoding sequences, required for early embryogenesis. To this end, we have performed a pilot large scale whole mount RNA in situ hybridization experiment to identify novel transcripts with tissue-specific expression pattern. We constructed two size selected plasmid cDNA libraries (inserts 1-2kb and >2kb, RNA from 1-20 somite embryos) and randomly sequenced approximately 200 clones. cDNAs that have novel sequences were used for RNA in situ hybridizations. Our results suggest that 5% of these clones have tissue specific expression patterns. To increase the probability of obtaining transcripts from a specific cell lineage, we generated transgenic zebrafish that express GFP in specific tissues. These fish are used to purify, by fluorescence activated cell sorting, the earliest lineage- specific progenitor cells from which RNA can be isolated for identifying lineage-specific transcripts. Given the availability of a large number of embryonic mutations and the ease of generating transgenic zebrafish, we believe that novel transcripts obtained from our search can be characterized in the context of both loss-of-function and gain-of-function. We have also developed zebrafish as a whole animal system to dissect the functions of non-coding sequences. By microinjecting DNA constructs that contain tissue specific promoters ligated to GFP, we demonstrated that functional cis-acting elements can be rapidly identified in living zebrafish embryos. We believe that this approach will allow us to identify trans-acting factors required for tissue specific expression of any developmentally regulated gene.

Shuo Lin
Institute of Molecular Medicine and Genetics
Medical College of Georgia
Augusta, Georgia 30912
telephone: 706-721-8762
fax: 706-721-8752
email: slin@mail.mcg.edu

Presentation format: Platform

Identification of ANF as a stroke-susceptibility gene by positional expression cloning using quantitative expression analysis

Michael P. McKenna, Gregory T. Went, Jonathan M. Rothberg, Stephen F. Kingsmore

Ischemic stroke is a common, complex disorder caused by a combination of genetic and environmental factors. A significant genetic component to stroke predisposition has been demonstrated by both rare Mendelian

inheritance, and by increased concordance in monozygotic compared with dizygotic twens. In addition, two recent studies have identified a major stroke-influencing quantitative trait locus. STR3, on rat chromosome 5, in the spontaneously hypertensive stroke-prone rat (SHRSP).

We report the use of Quantitative Expression Analysis (QEA) to compare the gene expression profiles in hearts from SHR and SHRSP rats fed a normal diet. QEA is a novel method that comprehensively and rapidly compares levels of gene expression between samples, with a limit of mRNA detection below 1 part in 125,000. QEA relies on uniform labeling and non-biased amplification of cDNA fragments, and uses information about specific sequences at the ends of the amplified products in conjunction with product lengths to assign electronically the potential genes that a given band represents (GeneCalling).

Using QEA, 12,000 fragments, derived from ~6000 genes, were compared in triplicate between heart mRNA samples of SHR and SHRSP rats. 29 differences (0.2%) of magnitude >1.5-fold were found. One gene shown by QEA to be expressed differently between SHR and SHRSP heart was atrial natriuretic factor (ANF), which maps within the segment of rat chromosome 5 containing STR3. Two fragments derived from ANF were expressed at 2-fold higher levels in SHRSP than SHR; abolition of these peaks with ANF-specific oligonucleotides confirmed the identities of these fragments (oligo poisoning). Sequence analysis of ANF from SHRSP and SHR rats revealed a substitution (G99S) that changes a highly-conserved glycine to serine residue, and that may influence peptide cleavage by the inactive prohormone. The finding that ANF is altered in expression and sequence between SHR and SHRSP rats, together with co-localization of ANF and STR3, suggests that an ANF mutation underlies STR3. Furthermore, the SHRSP allele is protective against stroke development, while the mutation observed in ANF in SHRSP rats is consistent with impaired function. Also in accord with this hypothesis is the increase in ANF expression in SHRSP heart, that may represent a consequence of a functional ANF impairment. The known role of ANF in control of vascular tone and intravascular volume, as well as the high density of ANF binding sites in the brain, are consistent with causality for STR3. Moreoever, increased ANF levels have been found in humans with acute stroke. Finally, these findings suggest the potential of the mutant ANF peptide as a preventative agent for stroke.

QEA, in combination with candidate gene mapping (positional expression cloning), offers a novel, rapid alternative to positional cloning, by identifying differentially expressed genes that map adjacent to disease loci. In common with conventional positional cloning, the comprehensive nature of QEA permits identification of disease loci without prior knowledge of pathophysiology. Positional expression cloning is anticipated to have broad applicability to animal models of inherited disease and, in particular, to multigenic disorders that defy conventional positional cloning strategies.

Richard A. Shimkets, Suresh G. Shenoy
CuraGen Corporation
555 Long Wharf Drive
New Haven, Connecticut 06511
email: skingsmo@curagen.com

Presentation format: Platform

EPODB: A bioinformatics system for the analysis of gene expression during erythropoiesis

A. Kel2, O. Kel2, N. Kolchanov 2, J. Schug1, C. Stoeckert3

1Center for Bioinformatics, University of Pennsylvania 2Institute of Cytology and Genetics, Novosibirsk 3Children's Hospital of Philadelphia

The next phase of the Human Genome Project entails genome scale, high-throughput generation of data leading to a deeper understanding of function. The management, analysis and visualization of data generated in this phase will undoubtedly be substantially more difficult than the sequence-oriented data that forms the foundation for the first phase of the Genome Project. EpoDB is a prototype system designed to explore the issues surrounding functional analysis of differentiation using vertebrate erythropoiesis as a model system. We will describe the current capabilities and tools developed in EpoDB for information capture, representation and visualization, and their use in the analysis of gene expression during erythropoiesis. EpoDB readily extends to other pathways in hematopoiesis and other differentiating systems.

C. Overton1, J. Haas1, F. Salas1,
1312 Blockley Hall (6021)
418 Guardian Drive
Philadelphia, Pennsylvania 19104-6145
telephone: 215-573-3105
fax: 215-573-3111
email: coverton@Wcbil.humgen.upenn.edu

Identification of consensus elements in 3'UTRs

For several years, we have used differential display- reverse transcriptase-PCR (dd-RT-PCR) to identify genes differentially expressed in response to environmental stresses. This process has provided us with hundreds of 3' UTR sequences. Recently, we have used a combination of theoretical and experimental approaches to identify consensus elements in these 3' UTRs. For many of these sequences (20-25 bp in length) we have performed electrophoretic mobility shift assays (EMSAs) to establish specific binding of proteins to the nucleotide sequence. For one such sequence (called C1) which is identical to an EST sequence, we set-up affinity columns, gel-purified the samples, and then obtained micro sequences. The extracted proteins were TopoisomeraseI, nucleolin (C23), and mucleo- plasmin (B23), all of which are DNA binding patterns. The reconigition sequences for these proteins have short sequences in the C1 consensus element, sporting the idea that the C1 consensus is a composite element with binding to three separate proteins. Similar approaches are being used to identify proteins to other regulatory elements.

Gayle E. Woloschak
Argonne National Laboratory
Center for Mechanistic Biology and Biotechnology
9700 South Cass Avenue
Argonne, Illinois 60439-4833
telephone: 630-252-3312
fax: 630-252-3387
email: woloschak@anl.gov

Presentation format: Platform

Automated annotation of genomic DNA sequence: The Genome Channel

Richard J. Mural, and the DOE Annotation Consortium.

Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA.

Automatic annotation of large amounts of genomic DNA sequence is, and will continue to be, a formidable challenge. Only by developing very efficient computational tools for the initial annotation of the sequence and then by treating these annotations as hypotheses and testing and verifying them in the laboratory will this problem be properly addressed. We are developing an engine which will provide a framework for the analysis and annotation of genomic DNA sequences. This system includes methods for data retrieval, visualization, data warehousing and data mining. The interface to this system is called The Genome Channel.

Results of the various analyses performed by the system are presented in this interface. A number of features including simple and complex repetitive DNA sequences, tRNAs, CpG islands, as well as the results of several gene finding programs, including a new version of GRAIL, which incorporates EST similarity in gene prediction and modeling, are included in the analysis. Links from the analysis to other data resources are also included. Currently the results from all the major human sequencing centers can be viewed in the Genome Channel.

Richard Mural
Life Sciences Division, Oak Ridge Natioanl Laboratory
1060 Commerce Park
Oak ridge, TN 37831
telephone: 423-576-2938
email: muralrj@ornl.gov

Presentation format: Platform

Genomic sequence analysis of a gene-dense region on Chromosome 21q22.3 and comparison with the existing transcript map: The emerging gene organization

The hunt for genes at a chromosome scale is expected to provide dense transcript maps spanning large DNA regions. One of the immediate goal is to provide a resource for scanning for candidate genes associated to genetic diseases, as a substitut e to traditional positional cloning. Human chromosome 21 has been used as a model for this approach, and more than 1,000 gene fragments are now mapped onto this chromosome. In parallel, genomic sequencing of the long arm of chr.21 has been initiated in a consortium of laboratories and is expected to be finished by the end of year 1999. The information provided by the sequence analysis has a unique value for

predicting gene organisation and assessing the previously assembled transcript maps. In particular, the distal part of chromosome 21 is of tremendous interest since it is extremely gene rich, and is associated to a number of genetic disorders, such as APECED disease. We are presenting here two examples of integrated gene search in 21q22.3: 1) analysis of (Abstract truncated during submission process)

Marie-Laure Yaspo
MPIMG
Ihnestrasse 73
Berlin
D-14195 GERMANY
telephone: 49-30-8413-1356
fax: 49-30-8413-1380
email: yaspo@mpimg-berlin-dahlem-mpg.de
url: http://chr21.rz-berlin.mpg.de

Presentation format: Platform