6th International Workshop
on the Identification of Transcribed Sequences

October 3-5, 1996 Edinburgh, Scotland

Table of Contents

Agenda

Wednesday October 2, 1996

5:30 pm - 7:30 pm -- Welcoming Reception
Edinburgh Hilton, Pavilion Restaurant

Thursday October 3, 1996

7:30 am - 8:15 am Continental Breakfast, Riverside Suite

8:15 am - Welcome/Opening Remarks

SESSION 1 - CHAIRS: Yun-Fai Chris Lau and Miriam Meisler

8:30 am - 8:50 am - James W. Fickett
Transcriptional context from DNA sequence: A beginning

8:50 am - 9:10 am - Marcelo Bento Soares
Subtractive hybridization strategies to facilitate gene discovery; Towards the development of libraries

Sponsors

- Cancer Research Campaign
- European Commission
- HUGO Europe
- Lothian and Edinburgh Enterprise
- U.S. Department of Energy
- Wellcome Trust

Corporate Contributors

Amgen
A global biotechnlogy company that discovers, develops, manufactures, and markets human therapeutics based on advanced cellular and molecular biology.

CIBA-Geigy
CIBA-Geigy are a leading worldwide biological and chemical group, based

enriched for full-length cDNAs

9:10 am - 9:30 am - J. Gregor Sutcliffe
A digital technology for analyzing expression of nearly all genes

9:30 am - 9:50 am - Kym Boycott
Isolation and incorporation of 44 new EST's into a physical map of the human X chromosome at Xp11.23-p11.22

9:50 am - 10:10 am - Lucy R. Osborne
Identification of genes deleted in Williams syndrome at 7q11.23 using direct cDNA selection and genomic DNA sequencing: A model for rapid gene discovery

10:10 am - 10:30 am - BREAK

10:30 am - 10:50 am - Katheleen Gardiner
Chromosome 21 - After transcriptional mapping: expression patterns and isochore boundaries

10:50 am - 11:10 am - Ian N. Hampson
Directional random oligonucleotide primed (DROP) global amplification of cDNA: Its application to subtractive cDNA cloning

11:10 am - 11:30 am - Sally H. Cross
Isolation of CpG islands and their use to isolate genes

11:30 am - 11:50 am -Steven J.M. Jones
Gene prediction within the C. elegans genomic sequencing project

12:00 pm - 1:00 pm LUNCH - PAVILION RESTAURANT

1:00 pm - 2:00 pm - **Poster Session I** - Ravelston Room

SESSION 2 - CHAIRS: James Fickett and Greg Elgar

2:00 pm - 2:20 pm - Sherman M. Weissman
cDNA analysis: (1) Complexities in deducing mRNA structure from genomic sequences, (2) Analysis of gene expression in small numbers of cells

2:20 pm - 2:40 pm - Michael Clinton
Differential display analysis of gene expression in the developing gonads of the chick embryo

2:40 pm - 3:00 pm - Miles B. Brennan
Systematic CRE/LOX deletion mutagenesis in the mouse

3:00 pm - 3:20 pm - Laurent Duret

in Switzerland, dedicated to satisfying needs in healthcare, agriculture and industry with innovative value-adding products and services.

Life Technologies
Producers of Gibco BRL brand names, Life Technologies supplies life science researchers and biomedical manufacturers with the special tools needed to do their important worki.

Pfizer, Inc.
Pfizer, Inc. is a research-based, global health care company. Their mission is to discover and develop innovative, value-added products that improve the quality of life of people around the world and help them ejoy longer, leatheir and more productive lives.

Zeneca Pharmaceuticals
Zeneca Pharmaceuticals is a leading international bioscience group active in three main areas of business - pharmaceuticals, agrochemicals and specialities - which provide products and services to improve human health, nutrition and quality of life around the world.

Searching for new regulatory elements: Large scale comparative analysis of metazoan non-coding sequences

3:20 pm - 3:40 pm - Esther E. Schmidt
Isolation and cloning genes involved in the oncogenesis of human astrocytic tumours by means of differential display

3:40 pm - 4:00 pm - BREAK

4:00 pm - 4:20 pm - John N. Feder
Transcript map surrounding the hemochromatosis candidate gene, HLA-H reveals several non-MHC gene families

4:20 pm - 4:40 pm - Catherine Nguyen
Quantitative analysis of differential expression between normal and deficient thymus

4:40 pm - 5:10 pm - Kevin G. Becker
Use of cDNA microarrays for the quantitative analysis of gene expression

5:10 pm - 5:30 pm - Richard J. Mural
The challenge of automated annotation of DNA sequences

Friday October 4, 1996

7:30 am - 8:30 am - Continental Breakfast, Riverside Suite

SESSION 3 - CHAIRS: Bento Soares and Laurent Duret

8:30 am - 8:50 am - Stephen W. Scherer
Gene identification on human chromosome 7q

8:50 am - 9:10 am - Gayle E. Woloschak
Use of a novel consensus sequence for regulon mapping

9:10 am - 9:30 am - Michael McClelland
The Cot effect and other quantitative PCR issues

9:30 am - 9:50 am - Esther van de Vosse
Improvements to cosmid-based exon trapping

9:50 am - 10:10 am - Wai-Choi Leung
Functional correlation of architectural elements on mRNA

10:10 am - 10:30 am -BREAK

10:30 am - 10:50 am - Gad Yagil
On the mechanism of DNA unwinding in yeast control regions

10:50 am - 11:10 am - Douglas Montgomery
In silico cloning of new transcripts using public databases

11:10 am - 11:30 am - Bernhard Korn
Construction and analysis of a gridded full length cDNA library generated from human fetal brain

11:30 am - 11:50 am - Greg Elgar
Comparative analysis with the Puffer fish, Fugu rubripes

12:00 pm - 1:00 pm - LUNCH - PAVILION RESTAURANT

1:00 pm - 2:00 pm - **POSTER SESSION II** - Ravelston Room

SESSION 4 -CHAIRS: Guiseppe Borsani and Catherine Nguyen

2:00 pm - 2:20 pm - Martin Ringwald
Gene expression information resource for mouse development

2:20 pm - 2:40 pm - Stephen F. Kingsmore
Positional cloning of the homologous beige mouse and human Chediak-Higashi Syndrome genes

2:40 pm - 3:10 pm - Donna Albertson
Characterization of predicted genes from the C. elegans genome sequence by visualization of expression patterns in whole animals by FISH

3:10 pm - 3:30 pm - Pierre Colas
New ways of charting and exploring regulatory networks

3:30 pm - 3:50 pm - Anthony J. Brookes
The genomic route to the isolation of novel genes

3:50 pm - 4:10 pm - BREAK

4:10 pm - 4:30 pm - Marie-Laure Yaspo
Transcriptional map of human chromosome 21: How far from completion?

4:30 pm - 4:50 pm - Patrick Onyango
Evaluation of candidate tumour suppressor genes from the chromosome 1p36 region: A locus commonly deleted in several cancers

4:50 pm - 5:10 pm - Marie-Josephe Pebusque
Progress in the generation of an integrated physical and transcription map of human 8p11-p21 region involved in breast cancers and hematological malignancies

5:10 pm - 5:30 pm - Jeffrey W. Touchman
2,006 ESTs from human chromosome 7-enriched direct cDNA selection libraries

5:30 pm - 5:50 pm - Petra Ross-Macdonald
Large-scale functional analysis of the S. cerevisiae genome

GALA DINNER --7:30 PM - PAVILION RESTAURANT

Saturday October 5, 1996

7:30 am - 8:30 am - Continental Breakfast - Riverside Suite

SESSION 5 - CHAIRS: Donna Albertson and Bernhard Korn

8:30 am - 8:50 am - Alla Rynditch
Novel characteristics of the 3q21 leukemia breakpoint region

8:50 am - 9:10 am - Guiseppe Borsani
Drosophila-related expressed sequences (DRES): A source of candidate genes for human diseases

9:10 am - 9:30 am - Yun-Fai Chris Lau
En masse terminal exon trapping of the human Y chromosome

9:30 am - 9:50 am - Miriam Meisler
Transcript mapping in the nonrecombinant interval for the mouse neuromuscular disease gene mnd2 and the corresponding region of human chromosome 2p13

9:50 am - 10:10 am - Pierre Pontarotti
New genes within the MHC: Cloning by paralogy: The case of the B30-gene family

10:10 am - 10:30 am -BREAK

10:30 am - 10:50 am - Masahiko Shiraishi
The isolation of CpG islands from human chromosomal regions 11q13 and Xp22 on the basis of thermal stability of DNA fragments covering these regions

10:50 am - 11:10 am - Nitin S. Udar
A gene rich region around the ATM locus - Physical map and identification of two new transcripts in the 11q23.1 region

11:10 am - 11:30 am - Jacob Odeberg
Analysis of differentially expressed genes using solid- phase approaches in transcript selection and in characterisation of cognate proteins expressed in bacterial systems

11:30 am - 11:50 am - Jennifer Phillips
Continuing studies on the subcellular localization of mRNAs in neurons

11:50 am - 12:10 pm - Thomas Werner
Successful modelling of the functional organization of transcriptional units

12:30 pm - 1:30 pm - LUNCH - PAVILION RESTAURANT

SPECIAL SESSION

1:30 pm - 1:50 pm - Canadian Genome Project

1:50 pm - 2:10 pm - Edinburgh Scottish Collaborative Centre

AFTERNOON SESSION

2:15 pm - 5:00 pm - DISCUSSION GROUPS AND MEETING SUMMARY

5:00 pm - Meeting Adjourned

SESSION 1

1. **Transcriptional context from DNA sequence: A beginning**

James W. Fickett
SmithKline Beecham Pharmaceuticals
King of Prussia, PA, USA

In the mapping of disease genes, in organism-wide gene inventories, and in drug discovery, a major challenge is to assign function to newly discovered genes. One important determinant of function is transcriptional context, and a number of experimental techniques are being developed and exploited to inventory the rough expression level of many genes in many particular tissues and disease states.

As genomic sequence accumulates, the analysis of transcriptional regulatory regions in the DNA will increasingly be able to contribute to the characterization of transcriptional context, and perhaps to the fundamental understanding of transcriptional regulation as well. Our goal is to write computer programs that take as input windows of DNA sequence and produce as output the particular conditions (possibly none) under which each window functions as a transcriptional activator or repressor.

Progress will be reported on the development of a prototype algorithm for the recognition of myotube-specific promoters and enhancers. This will include a discussion of a specialized data collection, characterization of the DNA-binding specificity of a number of transcription factors, co-occurrence patterns for multiple transcription factors, and a Bayesian scheme for choosing between the hypotheses of (1) myotube specific regulatory region, (2) other regulatory region, or (3) non-regulatory region.

2. **Subtractive hybridization strategies to facilitate gene discovery**

Maria de Fatima Bonaldo(1), Greg Lennon(2) and Marcelo Bento Soares(1,2)
(1)College of Physicians and Surgeons of Columbia University, New York, USA (2)The New York State Psychiatric Institute, New York, USA and (3)Lawrence Livermore National Laboratory, Berkeley, CA, USA

Large-scale single-pass sequencing of cDNA clones randomly picked from libraries has proven quite powerful to identify genes and the use of normalized libraries in which the frequency of all cDNAs is within a narrow range has been shown to expedite the process by minimizing the redundant identification of the most prevalent mRNAs. However, given the large scale nature of the ongoing sequencing efforts and the fact that a significant fraction of the human genes has been identified already, the discovery of novel cDNAs is becoming increasingly more challenging. In an effort to expedite this process as we strive towards the ultimate goal of identifying the majority of the human genes, we have developed and applied subtractive hybridization strategies to eliminate pools of sequenced cDNAs from libraries yet to be surveyed. Briefly, single-stranded DNA obtained from pools of arrayed and sequence I.M.A.G.E. clones are used as templates for PCR amplification of cDNA inserts with flanking T7 and T3 primers. PCR amplification products are then used as drivers in hybridizations with normalized libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) are purified by hydroxyapatite chromatography, converted to double-stranded circles and electroporated into bacteria. Preliminary characterization of a subtracted fetal liver-spleen library indicates that the procedure is effective to enhance the representation of novel cDNAs.

Towards the development of libraries enriched for full-length cDNAs

Kala Mayur(1), Maria de Fatima Bonaldo(1) and Marcelo Bento Soares(1,2)
(1)College of Physicians and Surgeons of Columbia University, New York, USA (2)The New York State Psychiatric Institute, New York, USA

It is anticipated that the existence of libraries enriched for full-length cDNAs will greatly facilitate positional cloning projects aimed at the identification of disease genes. In an effort towards this goal we developed and tested an approach for construction of libraries enriched for full-length cDNAs which involves: (i) pooling of mRNAs from different sources, (ii) size fractionation of the mRNAs, (iii) cDNA synthesis from each individual RNA size fraction, (iv) size selection of the resulting cDNAs according to the size fraction of the starting mRNAs, (v) separate cloning and propagation of each individual sub-library. In a feasibility study that we have conducted to assess the effectiveness of our procedure, we have generated 5 sub-libraries from a mixture of size fractionated mRNAs from human infant brain and placenta. Preliminary characterization of these libraries indicate that they contain cDNAs ranging in size from 0.35 kb up to 10 kb. Sequence analysis suggests that over 50% of the clones in each sub-library is either full-length or near full-length. Work is in progress to develop similar resources from a more comprehensive pool of mRNAs.

3. **A digital technology for analyzing expression of nearly all genes**

J.G. Sutcliffe(1), P.E. Foye(1), L. de Lecea(1), M. Carson(1), E. Thomas(1), B. Hilbush(2), K.W. Hasel(2)
(1)Dept. of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA (2)Digital Gene Technologies, Inc., La Jolla, CA, USA

We report a method called TOGA (TOtal Gene expression Analysis) that utilizes an 8-nucleotide sequence near the 3' ends of mRNAs, comprised of a 4-nucleotide restriction cleavage site and an adjacent 4-nucleotide parsing sequence, to give nearly every mRNA in the organism a distinct identity, regardless of whether the mRNA has been discovered previously. The restriction cleavage site defines the distance of the site from the 3' poly(A) tail for each mRNA. All 256 possible permutations of the adjacent 4 nucleotides are used one at a time as part of primer-binding sites in PCR-based assays performed on mRNA samples from tissue extracts to parse the expressed mRNAs into 256 pools, each of which is separated by electrophoresis for visualization of its product constituents. The 8-nucleotide sequence and measured length together provide a nearly unique digital address for each mRNA species. Data collected from manual and automated assays demonstrate that the appearance of a band generated by TOGA indicates the presence in the sample of an mRNA at a concentration proportional to band intensity. TOGA is reproducible, detects mRNAs of less than 0.001% prevalence, and with four iterations identifies and quantitates greater than 99% of the mRNAs expressed in a sample. Thus TOGA represents a powerful means for comparing mRNA expression profiles.

4. **Isolation and incorporation of 44 new ESTs into a physical map of the human X chromosome at Xp11.23-p11.22**

Kym Boycott
University of Calgary, Dept of Medical Genetics, Calgary, Alberta

Several genes have been cloned and mapped to the short arm of the human X chromosome between ZNF21 and DXS255 at Xp11.23-p11.22. These include the erythroid-specific transcription factor GATA1, the transcription factor for the enhancer E3 (TFE3), synaptophysin (SYP), and the genes for Wiscott-Aldrich syndrome (WASP), X-linked thrombocytopenia (WASP), Dent disease (CLCN5), and one form of synovial sarcoma. In addition, several new genes and expressed sequence tagged sites (ESTs) have been

identified in this region including MG21, MG44, MG61, MG81, DXS1011E, DXS7469E, RNPL, IS2, IS3, IS4, IS7. The genes for both X-linked congenital stationary night blindness (CSNB1) and Island Eye disease (AIED) have been mapped to this region of Xp11. In an effort to positionally clone these two disease genes we have generated an extensive transcript map incorporating all known genes and ESTs, as well as 44 new ESTs, and imbedded it in a detailed physical map of this region.

We have previously constructed a 2 Mb physical contig in Xp11.23-11.22 consisting of yeast artificial chromosomes (YACs) and cosmid clones (Boycott et al., 1996). The contig was constructed from the content of 43 DNA markers (including 20 new sequence tagged sites (STSs) and 12 genes or expressed sequences) for 99 YACs and 130 cosmids. The contig spans the distal marker ZNF21 and extends proximally to include DXS255, with an average STS marker density of one every 50 kb. We have also mapped an additional five published ESTs on the physical contig and the relative order of previously known genes and expressed sequence tags (ESTs) in this region is predicted to be Xpter - ZNF21 - IS7 - DXS7465E (MG66) - DXS7927E (MG81) - RNPL, IS3, IS4, WASP, DXS1011E, DXS7465E (MG21) - DXS7466E (MG44) - GATA1 - IS2 - DXS7469E (Xp664) - TFE3 - SYP (DXS1007E) - CLCN5 - Xcen.

A minimal set of YACs and cosmids were used for the isolation of transcribed sequences by direct cDNA selection of retinal, brain, fetal brain, and placental cDNA sets. We have found direct selection to be 75% efficient and have mapped 44 new ESTs to the physical map of the region between ZNF21 and DXS255. We are currently isolating full length cDNAs in an effort to collapse the 44 ESTs into a limited set of genes. A detailed map incorporating physical, transcript, and genetic information will be valuable for identifying disease causing genes that map to this region of the X chromosome.

Boycott, K.M., Halley, G., Schlessinger, D., and N.T. Bech-Hansen. (1996). Genomics 33(3): 488.

5. **Identification of genes deleted in Williams syndrome at 7q11.23 using direct cDNA selection and genomic DNA sequencing: A model for rapid gene discovery**

Lucy R. Osborne(1,2), Stephen W. Scherer(1,2), Duane Martindale(3), Johanna Rommens(1,2), Ben Koop(3), and Lap-Chee Tsui(1,3)
(1)Department of Genetics, The Hospital for Sick Children, Toronto; (2)Department of Molecular and Medical Genetics, The University of Toronto, Ontario; (3)University of Victoria, British Columbia, Canada.

Williams syndrome (WS) is a multi-system developmental disorder caused by the deletion of contiguous genes at 7q11.23. Hemizygosity of the elastin (ELN) gene can account for the vascular and connective tissue abnormalities observed in WS patients, but the genes that contribute to features such as infantile hypercalcemia, dysmorphic facies, and mental retardation remain to be identified. Identification of all genes within the deleted region will be necessary for a full understanding of the underlying etiology of WS. As part of our human chromosome 7 mapping project, we are constructing a physical map of the WS region using a combination of cosmids, YACs, and PACs. Although well-characterized YAC contigs flanking the WS region have been established, YACs in the WS commonly deleted interval appear to be highly unstable. We have thus far isolated overlapping 'sequence-ready' cosmid and PAC clones covering approximately 1 Mb of DNA from within the deleted region.

Our goal is to generate a transcription map of the entire deleted region (estimated to be 1.5 Mb) using a combination of direct genomic sequencing and cDNA selection. A comprehensive description of genes

localized to this region should be useful in understanding the genotype-phenotype correlation of this syndrome. We initially tested this combination of gene identification strategies on 500 kb of DNA from within the WS deletion. Nine cosmids representing the minimal tiling path of 500 kb of DNA were individually sheared and subcloned into M13 sequencing vectors, and 150-200 random clones picked and sequenced by fluorescence-labeled primers. The sequencing data were assembled into contigs and searched for open-reading frames and matches to known genes and ESTs. In addition, the same cosmid tile was used for cDNA selection using a variety of cDNA pools, and the retrieved clones were sequenced and analyzed for homology to known genes or ESTs.

cDNA selection experiments revealed three putative genes within the 500 kb region studied. None of these transcription units had been identified previously, although one, which showed homology to the RRM/RNP group of RNA-binding proteins, had been isolated as an EST (HUMORF D26068). The other two units showed no similarity to any known gene but were used to isolate longer clones from a fetal brain cDNA library. Genomic sequencing isolated seven genes from within the region, including the RNA-binding protein. Three of these genes had been cloned previously (ELN, LIMK, and RFC2), one had homology to restin and two others corresponded to ESTs. All genes contained identifiable exons. In addition, a number of ESTs were identified whose translation products showed most similarity to prokaryotic proteins. These were most likely due to E. coli contamination of the cosmid sublibraries used for sequencing.

While it will be important to extend our contig to cover the entire commonly deleted region in WS in order to obtain a complete description of the genes for further analysis, the direct DNA sequencing strategy combined with searching for ESTs present in public databases appears to be particularly effective. It may become more robust than direct cDNA selection procedure for gene identification as the number of EST entries continues to increase. It is clear, however, that no one technique will identify all the genes within a given part of the genome. In light of these experiments we are now extending both the cDNA selection and genomic sequencing strategies to include gene identification in the remaining interval of DNA from within the common WS deletion.

6. **Chromosome 21 - after transcriptional mapping: expression patterns and isochore boundaries**

Katheleen Gardiner and Laurent Villard
Eleanor Roosevelt Institute, Denver, CO, USA

I. Expression patterns. Gene identification efforts on human chromosome 21 have been effective in isolating partial cDNAs for a significant percentage of the genes contained within it. Because the original biological motivation for this work was an understanding of the Down Syndrome phenotype, the next step is to determine the developmental expression patterns and the functional roles of these genes. To determine expression patterns, 230 non-overlapping, non-repetitive cDNAs have been arrayed and screened with mRNA from 4 stages of fetal pig (brain and body) and 2 ages of mouse brain. This has defined a set of cDNAs expressed in developing brain and has defined a smaller subset that show potentially interesting differential expression patterns. In a continuing search for sequence homologies that may assist in functional determinations, all cDNAs are periodically used in Blast searches. Recent results indicate that a significant proportion of even the non-rare messages are still not represented either by functional matches or in the EST databases.

II. Isochore boundaries. The mammalian genome is a mosaic of regions of varying base composition (the isochores). Previous transcriptional mapping efforts have made clear that the gene distribution on 21q correlates well with the patterns in base composition. Recently, Fukagawa et al. cloned the transition between an AT-rich region and a GC-rich region on human chromosome 6, and showed that the segment

had high homology to the PseudoAutsomal Boundary of the sex chromosomes. They termed this sequence PABL (PAB-like) and showed that similar sequences occur in many copies in the human genome. Because there are many boundaries between regions of differing base composition on chromosome 21, we screened a chromosome 21 cosmid library with a PABL sequence. We have identified numerous PABL containing clones, a subset of which have been sequenced and placed on the physical map. Interestingly, some appear to be expressed and their locations correspond well with predictions from the base compositional map. Potential structural or functional or organizational roles for such sequences will be of interest to explore.

7. **Directional random oligonucleotide primed (DROP) global amplification of cDNA: Its application to subtractive cDNA cloning**

Ian N. Hampson, Lynne Hampson and T. Michael Dexter
Paterson Institute of Cancer Research, Department of Experimental Haematology, Christie Hospital, Manchester, United Kingdom

We originally developed the method of Chemical Cross Linking Subtraction (CCLS) (1 and abstract by Joanne M. Walter et al). The main limitation to this approach is the availability of mRNA. We now describe a rapid method of global PCR amplification of cDNA such that the strand sense is maintained. The products of this process are random primed fragments which facilitates uniform PCR amplification of total cDNA. Directional incorporation of an RNA polymerase initiator/promoter sequence allows efficient synthesis of total sense RNA from this material and the use of a biotinylated primer permits the separation of single stranded cDNA. Isolation of these products from different cell types provides a renewable source of single stranded target cDNA and driver RNA from limited cell numbers and we demonstrate their use for CCLS subtractive cloning of differentially expressed cDNAs.

References Hampson I.N., Hampson L. (previously Pope), Cowling G.J. and Decter T.M., Nucl. Acids Res. 20, pp 2899, 1992

8. **Isolation of CpG islands and their use to isolate genes**

Sally H. Cross, Victoria H. Clark and Adrian P. Bird
Institute of Cell and Molecular Biology, Edinburgh University, Edinburgh, Scotland

One of the major challenges in gene discovery, when using a positional cloning approach, is the identification of transcripts in genomic clones. Once the region in which a gene of interest resides has been narrowed down to a few kilobases a variety of techniques are employed to identify genes, for example exon trapping. An approach which we are developing involves the isolation of intact CpG islands from genomic clones which can then be used as probes to pull out full-length transcripts from cDNA libraries. CpG islands are short genomic regions of about 1 kb which are GC-rich and unmethylated. There are about 45,000 CpG islands in the human genome which mark and overlap the 5 end of 60% of genes. Using a methyl-CpG binding column we have exploited the unusual base composition of CpG islands in order to purify them from both genomic and cloned DNAs. To date we have constructed whole genome CpG island libraries from human, mouse and chicken. In addition we have single chromosome CpG island libraries from human chromosomes 22 and 18. One important feature of CpG island libraries is that the representation of each CpG island is independent of the expression pattern of the associated gene. Furthermore CpG islands are usually single-copy and therefore make ideal probes. We have also isolated CpG islands from a human cosmid clone and are currently applying the technique to a human PAC clone.

CpG islands are important genomic regions because they contain both promoter and transcribed parts of genes. Because of their high GC-content they are likely to be the regions which are most difficult to

sequence in the human genome sequencing project. Therefore although CpG islands are not associated with all genes their isolation provides an important resource for genome projects.

9. **Gene prediction within the C. elegans genomic sequencing project**

Steven J.M. Jones and Richard Durbin
The Sanger Centre, Wellcome Tust Genome Campus, Cambridge, United Kingdom

The C. elegans genomic sequencing project has so far derived over 46MB of sequence from the 100MB genome. Currently over 7,300 protein coding genes have been predicted and over 330 tRNA genes. The total number of protein coding genes within C. elegans is estimated to be in the range of 14-15,000. Approximately 50% of the predicted proteins show similarity to previously determined protein sequences in the public databases.

Protein coding genes are initially predicted using GENE FINDER (P. Green and L. Hillier, unpublished). GENE FINDER utilizes known codon biases in C. elegans coding regions, splice sites, start and stop signals to identify putative exons and dynamically attempts to predict a set of non-overlapping genes for a given sequence. Additionally, Hexamer (or diamino) frequencies are also now being used to identify putative protein coding sequences.

The initial GENE FINDER predictions are consolidated with C. elegans EST and protein homology data. BLASTN (Altshul et al. 1990) in conjunction with MSPcrunch (Sonnhammer and Durbin 1994) is used to map ESTs to the genomic sequence and confirm exonic sequences. ESTs also represent the only way that alternate splice variants can currently be detected in previously uncharacterised genes. Currently around 31% of the predicted genes possess one or more ESTs sequences. Exons are also detected by comparing the C. elegans genomic sequence to previously determined protein sequences using BLASTX and MSPcrunch. tRNA genes are predicted using the TRNASCAN method of Fichant and Burks (1991).

The database ACEDB (R. Durbin and J. Thierry-Mieg, unpublished) provides the environment within which all the gene appraisal, editing and annotation is carried out.

Altshul et al. (1990) J. Mol. Biol. 215:403-410. Fichant G.A. and Burks C.J. (1991) J. Mol. Biol. 220:659-671. Sonnhamer E.L. and Durbin R. (1994) Comput. Appl. Biosci. 10(3):301-307.

SESSION 2

1. **cDNA analysis: (1) Complexities in deducing mRNA structure from genomic sequences, (2) Analysis of gene expression in small numbers of cells**

S. Nalabolu(1), N. Baskaran(1), Y.C. Liu(1), Y. Prashar(2), K. Prakash(2), and S. Weissman(1).
(1)Yale University School of Medicine, Dept. of Genetics, New Haven, CT, USA (2)Gene Logic Inc. Columbia, MD, USA

Comparison of full length cDNAs and genomic sequences in the human MHC shows both the value of having genomic sequences and certain limitations in present methods for using genomic sequences to predict the structure of mRNAs and proteins. Problems that arise in identifying and characterizing mRNAs based on genomic sequences include the expected examples of missing small exons and misinterpreting intronless genes. In addition, several curious situations have been noted including the presence of large introns containing multiple internal transcribed sequences, the presence of a gene, B144, that encodes an approximately 700 base pair mRNA which occurs in as many as thirty alternatively spliced forms, the use of a non-canonical 5' splice site, and the detection at the protein level of a product whose apparent molecular mass is about twice that predicted from the nucleotide sequence.

In other work, we continue to adapt modified PCR procedures for the production of nearly full length cDNAs from very small amounts of starting material, and to apply 3' end restriction fragment display methods to preformed cDNA libraries. The combination of these approaches makes it possible to compare patterns of gene expression from very small numbers of cells such as highly purified murine hematopoietic stem cells.

2. **Differential display analysis of gene expression in the developing gonads of the chick embryo**

Haines L, McBride D and Clinton M
Roslin Institute, Edinburgh, Scotland

In mammals, sex depends upon the inheritance of a heteromorphic pair of sex chromosomes, X and Y, with the dominant Y chromosome in males leading to the formation of testis. However, this mechanism of sex determination can not operate in all vertebrates, for example, in most avian species the female is the heterogametic sex (ZW) and the male the homogametic sex (ZZ). At present, it is unclear whether ovary development is the result of a dominant W chromosome or that the testis develop under a Z chromosome dosage mechanism. The gonadal primordia, the genital ridges appear in the chick embryo at day 4 of development and superficial morphological differences are apparent between males and females by day 8. Sex determination is thought to occur around day 6 of development.

In an attempt to identify genes involved in avian gonadal development, we have used Differential Display to compare the patterns of gene expression in the developing ovary and testis. Genital ridge/gonads were excised from male and female chick embryos between days 4 and 9 of development and RNA extracted from these tissues was used to perform differential display analyses. Banding patterns obtained from male and female samples were compared and bands exhibiting sexually dimorphic patterns or intensity changes during development were identified. Gel fragments corresponding to these bands were excised, and the

DNA eluted, re-amplified and cloned. Here we report the results obtained from five such clones. In order to confirm the expression patterns demonstrated by differential display analysis, Northern hybridization was performed using these clones as probes. For two of these clones, no signal was detectable by Northern hybridization while the remaining three clones demonstrated patterns of expression which replicated the patterns seen by differential display. Of these transcripts, three represent novel sequences, one represents a transcript with a well documented developmentally regulated pattern of expression and one represents the chick homologue of a known signal transduction molecule considered to be involved in cell morphology changes.

3. **Systematic CRE/LOX deletion mutagenesis in the mouse**

Miles B. Brennan(1), Brian Sauer(2), and Ute Hochgeschwender(3)
(1)Eleanor Roosevelt Institute, Denver, CO, USA; (2)Laboratory of Biochemistry and Metabolism, NIDDK, Bethesda, MD, USA (3)Unit on Molecular Genetics, NIMH, Bethesda, MD, USA

The systematic identification of transcribed sequences has yielded many intriguing novel genes; their analysis presents a new challenge. One approach to the functional analysis of novel genes is the introduction of knockout (null) alleles into the mouse germline and the testing of the resulting mutants for phenotypes. While it is impractical to knockout each gene individually, engineering large segmental deletions spanning defined chromosomal regions would allow the phenotypic testing of null alleles in a large number of genes simultaneously. We are using the Cre/lox recombination system to introduce such defined deletions into the mouse germline.

We introduced lox sites at both the HPRT locus (X chromosome) and the POMC (proopiomelanocortin) locus (chromosome 12) by homologous recombination in embryonic stem (ES) cells. ES cell lines with each disruption have successfully colonized the mouse germline. We are now introducing new lox sites at sequences on chromosome 12 and on the X chromosome at specific distances from HPRT or POMC loci. After introducing these into the mouse germline, recombination will be catalyzed by the Cre recombinase expressed specifically in one cell embryos by the Ad5Ella promoter.

Analyzing knockouts of contiguous genes simultaneously will facilitate the identification of genes producing experimentally tractable phenotypes, especially extreme and specific effects on development or in adult function. The systematic character of this mapping offers important advantages over random insertional mutagenesis and enhancer/promoter trap strategies.

4. **Searching for new regulatory elements: Large scale comparative analysis of metazoan non-coding sequences**

Laurent Duret
Medical Biochemistry Department, Geneva University Hospital, Geneva, Switzerland

Genome sequencing projects currently produce hundreds of Mb per year. The major task of bio-informaticians now is to try to identify functional elements within this huge amount of genomic sequences. Some functional elements (such as tRNA genes, protein-coding regions, etc.) can be identified by using sets of rules or statistical methods. These methods are extremely useful, but their prediction power is often well under 100%, and they are limited to a small set of well known functional elements. Another possibility to search for functional elements is to use comparative analysis: if a genomic region has remained highly conserved during evolution, then it means that it is subject to a strong selective pressure and hence that it is functional. This method has some limitations: it requires to have genomic sequences from different taxa, not too closely and not too distantly related, and some functional elements cannot be

identified by comparative analysis. However, this approach proves to be very efficient, and its main advantage is that it does not require any prior knowledge of the functional elements to be searched.

We have used this method in a large scale project of identification of regulatory elements within non-coding parts of protein genes. 145 Mb of non-coding or non-annotated sequences from different metazoan taxa (essentially vertebrates, insects and nematodes) were extracted from databases and compared between each other using a combination of BLAST and LFASTA, to search for evolutionary conserved elements. This study revealed the existence of hundreds of very long highly conserved regions (HCRs) in non-coding parts of genes : 70% identity or more over 50 to 2000 nt between species that diverged 300 to 550 million years ago. Such a conservation is unexpected because it concerns fragments that are much longer than the regulatory elements known to date. Studying HCRs distribution within genes showed that functional constraints are generally much stronger in 3'-non-coding regions than in promoters or introns. The 3'-HCRs are particularly A+T-rich and are always located in the transcribed untranslated regions of genes, which suggest that they are involved in post-transcriptional processes (mRNA export, localization, translation, or degradation). Various insights suggest that many of these 3'HCRs play an important role in mRNA intracytoplasmic localization.

Information on HCRs (sequences, alignments, annotations, bibliographic references) are compiled in a database (ACUTS: database of Ancient Conserved UnTranslated Sequences) that will be made available to the scientific community.We now analyse all these conserved elements to try to detect some common features that could be used to set up rules for the prediction of regulatory elements in new sequences.

5. **Isolating and cloning genes involved in the oncogenesis of human astrocytic tumours by means of differential display**

Esther E. Schmidt, Koichi Ichimura and V. Peter Collins
Ludwig Institute for Cancer Research and Institute for Oncology and Pathology, Karolinska Hospital, Stockholm, Sweden

Differential display was used as an approach to identify genes that play a role in the tumorigenesis of human astrocytic tumours. Glioblastoma is the commonest and most malignant form of these tumours. RNA was extracted from clearly separated areas within the same tumour specimen, containing tumour cells and normal cells, respectively, and was subjected to differential display. Among 10 differentially displayed fragments successfully cloned into a plasmid vector, a 151 bp fragment isolated from a glioblastoma patient detected a 6 kb transcript on a Northern blot which was present in the tumour cells only. It was expressed in an additional 9/18 glioblastomas but not in normal brain tissues. Based on the 151 bp sequence, a 1.9 kb cDNA fragment was isolated from adaptor-linked U118MG cDNA using an approach combining 5 RACE and long-distance PCR (MarathonTM cDNA Amplification, Clontech). The sequence identified thus far represents a novel sequence with no significant homology to any known sequence. The cloning and sequencing of the full-length cDNA is ongoing. This strategy of searching putative oncogenes or tumour suppressor genes opens up new possibilities for understanding the processes involved in oncogenesis of human tissues.

6. **Transcript map surrounding the hemochromatosis candidate gene, HLA-H reveals several non-MHC gene families**

J.N. Feder, D. A. Ruddy, V. K. Lee, G. S. Kronmal, G. A. Mintier, A. Fullan, F. A. Mapa, N. C. Meyer, A. Basava, L. Quintana, E. McClelland, R. Domingo, D. B. Loeb, W. Thomas, Z. Tsuchihashi, R. C. Schatzman and R. K. Wolff
Mercator Genetics, Menlo Park, CA, USA

In the process of positionally cloning a candidate gene responsible for hereditary hemochromatosis (HH) (Feder et al, (1996) Nature Genetics 13, 399-408), we constructed a 1 megabase transcription map in a region 4 megabases distal to the HLA-A locus and several novel gene families were identified. A combination of direct cDNA selection, exon-trapping and genomic-sample sequencing were used to isolate 150 expressed sequence fragments which were ordered into an EST content map. Probes from each of the EST contigs were used to screen cDNA libraries. Besides the novel MHC class 1-like HH candidate gene, HLA-H, a family of five butyrophilin related sequences termed BTF1-5 were found. A gene with strong homology to the 52kd Ro/SSA lupus and Sjogren's syndrome autoantigen which we called RoRet was also identified. Both, the BTF family and the RoRet genes share a exon of common evolutionary origin called B30-2 (Vernet et al., (1993) J. Mol. Evol. 37, 600-612) . This exon, which was originally isolated from the HLA class 1 region has "shuffled" into several genes along the chromosome distal to the MHC. Here, we provide more examples. Also identified in the contig were two genes with predicted structural similarity to a previously cloned type 1 sodium phosphate transport gene (Chong, S. S., et al., (1993) Genomics 18, 355-359). As a further measure to isolate all of the genes in the HH candidate region, we completely sequenced 250 kb of genomic DNA and cluster histone genes were identified. The region around the HH locus therefore appears to be rich in gene families and duplicated sequences.

7. **Quantitative analysis of differential expression between normal and deficient thymus**

Catherine Nguyen(1), Samuel Granjeaud(1), Magali Gualandi(2), Philippe Naquet(2) and Bertrand Jordan(1)
(1)Genome Structure and Immune Functions Group, (2)Lympho-stromal Cell Interactions During Terminal T Lymphocyte Differentiation Group, CIML Luminy, Marseille, France

The CD3-zeta chain of TCR-CD3 complex plays a pivotal role in the activation of T cell responses and in the selection of the T cell repertoire. In zeta knock out mice, the T cells have a profound reduction in the surface levels of TCR-CD3 complexes and these animals have poorly developed thymuses.

In order to find new genes differentially expressed between normal and zeta knock out thymus, we set up two hybridizations on the same set of 3,072 genes with complex probes made from total RNA of zeta knock out or wild type thymus. After quantitative measurement of the amount of hybridized probe on each colony, the intensity ratios zeta-ko/wild type are calculated for each. 171 cDNA clones were selected showing a significant stimulation or a repression in zeta-ko thymus. Additional hybridizations performed with complex probes made from RNA of different cell types (such as macrophage, thymoctye, epithelial cell line under different stimulation) or tissues (lymph node, spleen...) allow to precise the selection. The 38 clones representing the most significant profile were tag sequenced (HGMP Resource Centre, Hinxton). Among these, 5 are highly homologous to known genes, 27 are new or related to mouse EST or human EST. "new" clones are tested on Northerns to determine the size of the mRNA. Three of them have been analysed by tissue in situ hybridization and show selective transcription in certain cell types.

8. **Use of cDNA microarrays for the quantitative analysis of gene expression**

Bittner, M.L.(1), DeRisi, J.(2), Meltzer P.S.(1), Becker, K.G.(1), Penland, L.(2), Ray, M.(1), Su, Y.(1), Brown, P.(2), and Trent, J.M.(1)
(1)Laboratory of Cancer Genetics, National Center for Human Genome, Research, NIH, Bethesda, MD,

USA (2)Department of Biochemistry, Stanford University Medical Center, Stanford, CA, USA

(The NCHGR/Stanford groups contributed equally to this work.)

cDNA microarrays provide a powerful tool for the study of gene expression patterns associated with complex biological phenomena. The cDNA microarray system is based upon the robotic printing of cDNAs on glass slides, and simultaneous two-color fluorescence hybridization (Schena et al., Science 270:467). A high density array of 1,161 DNA sequences was used to search for differences in gene expression associated with tumor suppression in a melanoma cell line (UACC-903) and its chromosome-6 suppressed subline [UACC-903 (+6)] (Trent et al., Science 247:568). The microarray contained cDNAs from two principal sources: 1) 674 genes derived from the Unigene set of the 1NIB normalized cDNA library; and 2) 183 genes resulting from reciprocal subtractive hybridizations between cDNAs from the tumorigenic UACC-903 cell line and its suppressed derivative. Fluorescent probes were generated by reverse transcription of mRNA from UACC- 903 cell line and UACC-903 (+6). A second experiment underway will examine the expression patterns of a large set (approx. 150) of zinc finger genes derived from human brain. These genes are being analyzed for their relative developmental and tissue specific expression patterns. Results of these studies suggest the enormous potential for applying this approach to the study of human gene expression.

9. **The challenge of automated annotation of DNA sequences**

Richard J. Mural
Biology Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA E-mail: muralrj@ornl.gov

As the analysis of genomes moves into the large-scale sequencing phase, the identification and annotation of biologically important features in anonymous DNA sequences become an even larger challenge. There are currently no systems capable of providing uniform high quality annotation of the estimated 2 million bases of DNA sequence which, in the near future, will be generated daily. Nor does there seem to be any consensus of what should be included in the annotation of genomic sequence. Issues of data ownership, third party annotation and database curation further complicate the resolution of this problem. In addition to timely annotation it is imperative that annotation be continually updated particularly in light of the rapid rate of new data acquisition.

Some of the issues involved in addressing these problems and some of the tools that are being developed to provide timely analysis of new sequences will be discussed. Systems for providing a user a personalized view of DNA sequence data will be discussed as will tools for data-mining and complex querying of multiple data resources. Examples of the kinds of information which can be obtained from analysis of large regions of uncharacterized DNA sequence will be illustrated.

SESSION 3

1. **Gene identification on human chromosome 7q**

S.W. Scherer(1), J. M. Rommens(1,2), S. Soder(1), G. Traverso(1), J., McArthur-Morrison(1), I. Wing-Yuk Szeto(1), L. Osborne(1), E. Belloni(1), H.H.Q. Heng(1), D.W. Martindale(3), B.F. Koop(3) and L.-C. Tsui(1,2)
(1)Department of Genetics, Research Institute, The Hospital for Sick Children (2)Department of Molecular and Medical Genetics, University of Toronto, Toronto, Canada; (3)University of Victoria, British Columbia, Canada

Using a multi-level mapping approach we have constructed an integrated genetic and physical map of the long arm of human chromosome 7 (7q). The map was generated by grouping over 1600 DNA markers to 30 different intervals with the use of somatic cell hybrids and rearrangement breakpoints and then ordering these amongst 2,443 YAC clones, 2400 cosmids, and 310 PAC clones. We estimate that over 95% of 7q is now covered in contigs. This mapping strategy has allowed the incorporation of physical and genetic landmarks into a fully integrated map providing a molecular framework for large-scale gene identification and mapping, and positional cloning of disease genes. So far 170 known genes and over 200 ESTs that were previously mapped to chromosome 7 have been positioned on our map. In order to isolate and map the remainder of the estimated 2500 genes on chromosome 7q three strategies are being used: (1) we are continuing to place all the genes and ESTs in the public domain on our chromosome 7 map; (2) genomic DNA sequencing: our mapping study has identified at least 17 regions of 7q that are not represented in the YAC map. Subsequent analysis has indicated that the majority of the "gap" regions can be filled with cosmids and/or PACs and that these regions are usually gene rich. Therefore these clone contigs are also being used as high resolution sequence-ready maps for gene discovery. In addition, positional cloning experiments of disease gene loci including Williams Syndrome (7q11.23), NIDDM (7q21.3), SHFM1/EEC (7q21.3-q22.1), SLOS (7q32.1), XRCC2 (7q36.1), HPE3 (7q36) and cancer associated regions mapped to 7q22, 7q31.2 and 7q34 are underway. While the techniques of direct cDNA selection, exon-amplification, and searching for evolutionarily conserved sequences have yielded several candidate genes for these loci, we have found that genomic DNA sequencing is a more robust method of gene identification. For example, >500 kb of DNA commonly deleted in Williams-Beuren syndrome patients at q11.23 was sequenced (see abstract by Osborne et al.) and 6 transcribed sequences were found that were not identified using the other techniques. (3) In a global approach to isolate transcribed sequences from chromosome 7 we have performed direct cDNA selection experiments on tiling paths of contiguous cosmid clones spanning regions of 7q and a chromosome 7-specific cosmid library. To maximize the number of different chromosome 7 transcribed sequences retrieved we have limited the amount of genomic DNA in each selection experiment to < 3 Mb and then subjected this to two rounds of hybridization with primary cDNAs pools from 12 fetal and adult tissues. In each experiment >90% of the clones isolated mapped back to human chromosome 7. Already over 500 cDNA fragments have been sequenced and mapped back to defined positions on chromosome 7. Our data is released at regular intervals at: http://www.genet.sickkids.on.ca/chromosome7/. Supported by the Canadian Genome Analysis & Technology Program.

2. **Use of a novel consensus sequence for regulon mapping**

Gayle E. Woloschak, Tatjana Paunesku, Aleksandar Milosavljevic
Argonne National Laboratory, Center for Mechanistic Biology and Biotechnology, Argonne, IL, USA
CuraGen Corporation, 322 East Main Street, Branford, CT, USA

In the process of identifying genes differentially expressed in ultraviolet-exposed cells, we identified a transcript having a 25-base-pair region that is highly conserved among a variety of species, including Bacillus circulans, pumpkin, yeast, Drosophila, mouse, and man. When in the 5' region (flanking region or UTR) of a gene, the sequence is predominantly in +/+ orientation with respect to the coding DNA strand; while in coding region and the 3' region (UTR), the sequence is most frequently in the */+ orientation with respect to the coding DNA strand. In two genes, the element is split into two parts; however, in most cases, it is found only once but with a minimum of 11 consecutive nucleotides precisely depicting the original sequence. It is found in a large number of different genes with diverse functions (from human ras p21 to B. circulans chitosanase). Gel shift assays demonstrated the presence of a protein in HeLa cell extracts that binds to the sense and antisense single-stranded consensus oligomers, as well as to double-stranded oligonucleotide. Because of its location in transcripts, this consensus element can be used for mapping cDNAs that comprise the regulon governed by the element. Experiments are underway using high-density membranes to identify cDNAs with the sequence. This work was supported by the U.S. Department of Energy, Office of Health and Environmental Research, under Contract No. W-31-109-ENG-38 and NIH grant #ES-07141-02.

3. **The Cot effect and other quantitative PCR issues**

Francoise Mathieu-Daude, John Welsh, Thomas Vogt, Rhonda Honeycutt, Karen Evans, Frank Kullmann, and Michael McClelland
Sidney Kimmel Cancer Center, San Diego, CA, USA

RNA fingerprinting by Arbitrarily Primed Polymerase Chain Reaction (RAP-PCR), and its variant, Differential Display (DD), are effective tools for studying differential gene expression in a variety of experimental systems. Once a differentially regulated gene is identified, it is usually necessary to verify its expression pattern independently. Therefore, we have explored several novel quantitative PCR concepts.

Relative quantitation by low stringency PCR: When a change in the abundance of a specific message is of interest, relative quantitation can be achieved by low stringency PCR using primers directed toward the proper sequence, but encouraged to engage in limited mismatch priming. The protocol is similar to RNA fingerprinting, except that specific (rather than arbitrary) primers are used, and reactions are stopped after various cycle numbers, as in other quantitation protocols. Products from mismatch priming serve as internal controls for a variety of reaction parameters. Because there can be a great many of these mismatched primers, effects due to sequence peculiarities of the control molecules are averaged out.

The Cot effect: Curiously, as a quantitative PCR reaction proceeds, more abundant products accumulate disproportionately slowly, even before the plateau effect due to limiting reaction components is reached. The "Cot effect" appears to be due to progressively faster reannealing of more abundant products as they accumulate. Conversely, less abundant products accumulate disproportionately rapidly. In RNA fingerprinting, the consequence of the Cot effect is that quantitative differences in an arbitrarily primed product between samples can be erased if the Cot effect is encouraged (e.g. by too many cycles).

Target vs. standard titration: Many examples in the literature, an internal standard is titrated against a fixed amount of RNA, for the purpose of quantitating a specific message by PCR. This approach works, but requires repeated sampling of the reaction during the log-linear stage of the reaction. Another strategy is simpler. First, a calibration curve for several different standard concentrations against a titration of RNA concentrations is generated. Then, the unknown is titrated against a fixed amount of standard. When this is done, the concentration of the target molecule can be read directly from a standard curve, because some point on this curve represents the exact ratio and absolute amounts of the control and target RNAs. Conversely, when the calibration curve is generated by titrating the standard against a fixed amount of the

target RNA, the curve does not contain, except by wild chance, the exact ratio and absolute amounts of control and target RNAs.

4. **Improvements to cosmid-based exon trapping**

Nicole Datson, Esther van de Vosse, Paola van der Bent, Hans Dauwerse, Emile de Meijer, Joris Heus, Gert Jan van Ommen and Johan den Dunnen
MGC-Dept. of Human Genetics, Leiden University, The Netherlands

Exon trapping has become a recognized method for the efficient isolation of coding sequences from large genomic regions. One of the merits of exon trapping is its independency of gene expression by detecting coding sequences at the DNA level. Consequently, genes with a complex or low level of spatial or temporal expression are identified in addition to ubiquitously expressed genes. With the complete sequence of the human genome coming within reach, computer prediction will become an important tool to identify new genes. Since some 70-90% of genes can be expected not to be represented by an EST, exon trapping will become an efficient system for in vivo confirmation of predicted genes. We have recently described a series of large-insert cosmid-based exon trap vectors, the sCOGH-vectors, which allow the scanning of entire cosmid inserts in one go for the presence of exonic gene fragments (1). Using this cosmid-based system we have successfully isolated multi-exon spliced products containing up to 7 contiguous exons of the dystrophin gene and 12 of the PKD1-gene. Several new variants have been constructed, including a YAC fragmentation vector. The latter facilitates the systematic and directive scanning of a YAC fragmented at Alu or LINE repeats for exons. The trapping of multiple exons of a gene in a single spliced product reduces the labour intensive analysis caused by repeated isolation of segments of the same gene. However, amplification of the longer RT-PCR products offers a technical challenge intrinsic to the system. We have introduced several changes in the protocol to improve RT-PCR yields, including long-range RT-PCR conditions, the use of polyA-selected RNA, new primers lowering the disturbing effects of background DNA, and a one tube RT-PCR system. Currently, we standardly use a 3'RACE to enable isolation of polyadenylated transcripts containing a 3' terminal exon, in addition to the RT-PCR directed at the trapping of internal exons. To facilitate simple determination of transfection efficiency we have inserted a Green Fluorescent Protein gene in one of the vector exons. This allows easy non-invasive monitoring by counting the number of fluorescent cells.

1. Datson et al., NAR 24: 1105-1111 (1996).

5. **Functional correlation of architectural elements on mRNA**

Wai-Choi Leung(1,2), Takashi Kishimoto(1,2), Calvin H. Leung(1,2), Shaoxiong Chen(3), Linda Hyman(3), and Maria FKL Leung(1,2)
(1)Division of Molecular Pathology, Department of Pathology and Laboratory Medicine, (2)Tulane Cancer Center and (3)Department of Biochemistry Tulane University School of Medicine, New Orleans, LA, USA.

We have further refined our approach in defining the architecture of mRNA. Briefly, the entire nucleotide sequence of a mRNA is folded into a predicted optimal structure using an energy minimization algorithm. Segment Analysis is then performed on the folded RNA by tracing the RNA polynucleotide backbone in a-tracings and b-tracings which revealed both short range and long range interactions. A double stranded region which appears in either a- or b-tracings is a Closed Region formed by short range interaction. The RNA sequence in a Closed Region can be independently refolded into a structural element identical to the corresponding structure on the folded entire RNA. On the other hand, an Open Region is contributed by long range interaction. Its sequence is folded into an alternate structure different from the corresponding

region on the folded entire RNA.

An Energy Map can be constructed to describe the location, size, energy content and energy density of Closed Regions on a mRNA. The presence of a Closed Region can be experimentally verified by its ability to reduce specific yields in a modified RT-PCR assay based on our observation that the log specific yield of an amplified RNA segment is inversely proportional to the sum of free energy in a Closed Region.

This approach readily generates molecular models for mRNAs on which more refine structural analyses can be performed. On the other hand, these molecular models provide sufficient structural information for functional assignment of architectural elements:

1. RNA Translocation: In Drosophila, the dorsal-ventral axis is determined by translocation of RNA to specific locations in the egg. The 3'UTR of the translocated RNAs are thought to recognize an adapter protein linking the RNA to microtubular apparatus which effects migration of RNA. Our analysis of the bicoid mRNA indicated that the 3'UTR region formed an extended stem structure protruding from the body of the folded mRNA, thereby promoting binding with the adapter protein. Similar structures have also been observed in other translocated RNA involved in development of Drosophila, e.g. nanos, etc., and in Xenopus, e.g., Vg1. Interestingly, in gurken mRNA, the 5'UTR directs translocation and exhibited a similar structure.

2. Nuclear Export: In HIV, the export of genomic RNA to the cytoplasm is mediated by binding of Rev protein to the Rev Responsive Element (RRE). Our analysis demonstrated that the RRE element exhibits an extended architectural element protruding from the viral genomic RNA. We hypothesized that this structure would promote the binding of Rev protein and facilitate export of viral RNA. Similarly structure is also observed for the cis-repressive element (CRS) which has previously been defined by genetic analysis as an additional binding site for Rev protein to facilitate nuclear export of viral genome.

3. Transcription Termination: Mutagenesis of the adh2 gene of Saccharomyces cerevisiae in a reporter construct indicated that the recognition site for transcription termination resided in a 19 nucleotide stem structure bearing two single nucleotide bulges. Structural analyses of the adh2 RNA model indicated that this stem bears an energy density of about -0.3 kcal/mol/bp. This RNA stem folded into an alternate structure in mutants which lost the termination function and was regenerated in a compensatory mutant which regained two thirds of termination function.

4. Translational Control: We observed that in a number of mRNAs known to exhibit translational control, the extreme 5' end sequence formed a stem structure with total energy value exceeding -50 kcal/mol. These elements also bear an energy density of -0.8 kcal/mol/bp or more, indicating high stability. Comparison of the stem structures for mRNA under similar translational control, e.g., p53 and CDK 4 mRNAs, revealed common sequence as potential recognition site for RNA-binding protein.

6. **On the mechanism of DNA unwinding in yeast control regions**

Gad Yagil
Department of Cell Biology, The Weizmann Institute, Rehovot, Israel; EBI, Cambridge, United Kingdom

The formation of an unwound DNA region is an essential step in gene transcription. The size and state of the unwound regions are nevertheless far from clear. A number of experimental tools for the characterization of unwound DNA are currently available, including: a. Single strand specific nucleases (S1, P1, mung bean nucleases), which serve to identify unwound regions. b. Two dimensional topoisomer analysis, which serves to determine the extent of unwinding. c. Conformation specific DNA reagents

which can distinguish between the various paranemic structures unwound DNA can assume. The paranemic structures include cruciforms, H- form DNA, B-Z junctions, paranemic duplexes and strand separated DNA (cf. Yagil, Crit. Revs. Bioch. and Mol. Biol., 26: 475-559, 1991).

The three techniques mentioned were applied to two strong yeast promoter regions containing long pyrimidine tracts, inserted into negatively supercoiled plasmids. Pyrimidine tracts are abundant in yeast promoter regions (Yagil, Yeast 10: 603) and are known to be attacked by single strand specific nucleases in eukaryotic nuclei. The principal P1 cleavage points of the DED1 promoter region maps within a sequence containing a pyrimidine tract of 40 bases; 2d topoisomer analysis indicates the unwinding of 4 primary turns. In the promoter of CYC1, a region of 33/36 pyrimidines is the principal P1 sensitive region. The limited symmetry of the P1 sensitive regions favors the formation of a paranemic duplex in the unwound region. Previous work showed that the chicken beta globin promoter and SV40 control regions assume an unwound state by 2d topoisomer analysis. It is proposed that polypyrimidine tracts serve as unwinding centers for DNA in eukaryotic genes, their length determining the extent of gene transcription.

To further examine this possibility, a program designed to list and report the frequency of binary homotracts in sequenced DNA databases was written. Application of the program, TRACTS, to a large selection of eukaryotic sequences, led to the conclusion that all purine or all pyrimidine tracts (R.Y tracts) are highly overrepresented in almost all eukaryotic genomes. Organelle genomes (mitochondria and chloroplasts) show a similar overrepresentation. Tracts which are all G,T or all A,C (K.M tracts) are overrepresented to a nearly similar degree, while A,T or G,C rich tracts are only marginally overrepresented. In the promoter regions of sequenced yeast chromosomes R.Y tracts longer than 15 nt are present at an nearly 50 fold excess over random DNA. This further strengthens the possibility that the R.Y tracts have a role in gene regulation.

7. **In silico cloning of new transcripts using public databases**

P. Sanseau, R.W. Gill, D.S. Montgomery, M.D. Oxer, S. Taylor, I.J. Purvis and C.W. Dykes
Glaxo-Wellcome Medicines Research Centre, Genomics and Advanced Technology & Informatics Research Units, Gunnels Wood Road, Stevenage, Herts, United Kingdom

With the continuous growth of the public databases, in silico cloning is becoming a method of choice to obtain new genes of interest. The use of the information available in databases such as dbEST and its minimally redundant version (Unigene) has already significantly decreased the time and resources required to clone target genes. In theory by using the information one should be able to perform electronic northerns, zooblots, cloning genes or extend known or unknown gene families from one's desk-top computer. Moreover the chromosomal mapping of thousands of human ESTs using the radiation hybrid panels will help positional cloning projects. Rapid in silico cloning is only possible if bioinformatics tools are available. Therefore new web interfaces have been developed in house to rapidly obtain new genes and analyse the corresponding sequences for characteristics such as potential open reading frames. We will describe the results of our attempts to in silico clone a number of current target genes and extend gene families out of dbEST or other public databases. Finally we will indicate how in silico cloning information points the way to further "wet" experimentation.

8. **Construction and analysis of a gridded full length cDNA library generated from human fetal brain**

Stefan Wiemann, Bernhard Korn, and Annemarie Poustka
Molecular Genome Analysis, German Cancer Research Center, Heidelberg, Germany

We have constructed a cDNA library from human fetal brain that contains 70-80 % full length

representations of primary transcripts. Single stranded cDNA was generated from fetal brain poly A+ RNA using an oligo dT primer, the =84cap switch oligonucleotide=93, and a reverse transcriptase deficient of RNase activity. Limited amplification of primary cDNA was performed under long range conditions using a combination of Amplitaq and Pfu DNA polymerases to minimize errors during PCR and to allow for amplification and subsequent cloning also of long cDNAs. The cDNA was cloned directionally into plasmid pAMP1 using the uracil/glycosylase cloning system. The library consists of 120,000 plasmid clones. The average insert size is 1.8 kb due to the selection for short inserts inherent to cloning in a plasmid vector. To date, 3,800 independent clones have been picked in 384 well microtiter plates and spotted on nylon membranes for hybridization analysis. No contaminating clones harboring copy rRNA were detectable in the array of 3,800 clones. As expected, 2 % of all clones were positive with a beta actin probe. The insert size is 1.9 kb in >80 % of these clones, reaching from the poly A tail to the described transcription initiation and cap addition site of the gene. Random sequencing of 100 clones revealed 49 known human genes, 12 homologous genes from human or other species, 26 genes that had hits only in EST databases and 13 novel genes with no hits in any database. 75% of the known genes were full length, most of them extending the published sequences at their 5=92-ends. Analysis of sequences upstream the first ATG revealed a high GC content (64 %) to be present in most cDNAs with high occurrence of CpG dinucleotides. In most novel cDNA clones the first ATG meets the consensus criteria of translation initiation suggested by M. Kozak. Currently, construction of cDNA libraries from other tissues, and testing of other cloning vectors for the selective cloning of longer (> 4 kb) cDNAs is underway.

9. **Comparative analysis with the Puffer fish, Fugu rubripes**

Greg Elgar
Molecular Genetics, Department of Medicine and UK HGMP Resource Centre, Cambridge, United Kingdom

The Japanese Pufferfish, Fugu rubripes, has a haploid genome of 400Mbases. It contains a similar gene set to mammals with the consequence that gene density is much higher. Many characteristics of genes are shared between mammals and teleost fish including high sequence similarity and conserved gene structure. Generally however, as well as being more densely spaced, Fugu genes are much smaller than their mammalian counterparts. This makes working with the Fugu genome much simpler than working with mammalian genomes. The Fugu genome has been used as a comparative tool in the analysis of gene structure, identification of regulatory elements, identification and confirmation of coding sequences and in whole genome analyses.There is some evidence that Fugu shares a degree of conserved synteny with mammals, a fact that has raised a great deal of interest with a view to positional cloning projects. However, little is known of the extent of these syntenies. The MRC is currently funding the Fugu Landmark Mapping Project. The aim is to sequence scan 1000 cosmids from a well characterized and publicly available genomic library which will provide a resource in a number of different areas including gene identification.

Each cosmid is analysed for a number of different features and the data presented on the world wide web interface. To date nearly 200 cosmids have been sequenced and analysed. A summary of data from this project will be presented as well as some specific examples of the ways in which the project can aid in gene identification and characterization. Because of the likelihood of finding more than one Fugu gene on a single cosmid clone, some physical linkage data is also available through this project.

SESSION 4

1. **Gene expression information resource for mouse development**

M. Ringwald(1), R. Baldock(2), J. Bard(3), D. Begley(1), G. Davis(1), D. Davidson(2), J.T. Eppig(1), M. Kaufman(3), J. Richardson(1), and L. Taylor(1).
(1)The Jackson Laboratory, Bar Harbor, ME,USA; (2)Human Genetics Unit, Edinburgh, Scotland; (3)Edinburgh University, Edinburgh, Scotland

As genome research shifts from the identification of genes to understanding their function, the amount of expression data, and the scale in which these data are generated grow at an enormous pace. Expression data are complex. Different types of expression assays give different but complementary insights into expression patterns, and new methods for measuring gene expression are developed rapidly. We urgently need tools that enable integrated storage and analysis of gene expression information.

We are developing a database of gene expression information for the laboratory mouse. The database is designed to store primary data from the different expression assays. In this format, the data can be integrated, new data and new assay types can be added, and novel insights resulting from new data can be represented. The system should enable the community to put bits and pieces of the various types of expression data together, and thus to successively gain knowledge about what products are made from a given gene, and when and where these products are expressed. Expression patterns are described using a standardized anatomical dictionary that models the anatomy hierarchically (from body region to tissue to tissue substructure) to allow for continuous refinement of the nomenclature system, to fit with the different resolution of analysis methods, and to facilitate user annotation of expression data. For in situ studies, the textual annotations are complemented with digitized images of original expression data. 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 interconnection with other databases (e.g. sequence databases, databases for other species) will place the gene expression data into the appropriate biological and analytical context.

We have implemented a Gene Expression Database prototype in the Sybase relational database system and have built interfaces that allow detailed standardized descriptions of expression data, including indexing and interactive labeling of 2D images, and electronic submission of these data from research laboratories. The submission system provides on-line links to other databases and electronic validation tools to facilitate data annotations and cross-referencing. Prototype interfaces are refined based on our experience in annotating expression data from the literature and on feedback from test laboratories, with the aim to develop a user friendly system for the community at large.

An index of expression citations from newly published research reports documenting data on endogenous gene expression during mouse development has been established. The index is updated daily and includes reference(s), gene(s) and embryonic stage(s) analyzed and expression assay(s) used. A searchable version of the Gene Expression Index is available at http://www.informatics.jax.org/gxd.html, or from our mirror site in the UK at http://mgd.hgmp.mrc.ac.uk/gxd.html.

2. **Positional cloning of the homologous beige mouse and human Chediak-Higashi Syndrome genes**

Maria DFS Barbosa(1), Velizar T. Tchernev(1), Jennifer A. Ashley(2), John C. Detter1, Quan A. Nguyen(1), Dipti Chotai(3), Charles Hodgman(3), Roberto CE Solari(3), Stephen J. Brandt(4), and Stephen F. Kingsmore(1)

(1)Dept of Med and Center for Mammalian Genetics, Univ of Florida, Gainesville, FL, USA (2)Dept of Biochemistry, Univ of Texas Southwestern Medical Center, Dallas, TX, USA (3)Cell Biology Unit, and Advanced Technologies and Informatics Unit, Glaxo Wellcome Medicines Research Center, Hertfordshire, UK (4)Department of Medicine, Vanderbilt University, Nashville, TN, USA

Vesicular transport to and from the lysosome and late endosome is defective in patients with Chediak-Higashi syndrome (CHS) and beige mice (bg). CHS and bg cells have giant perinuclear vesicles with characteristics of late endosomes and lysosomes that arise from dysregulated homotypic fusion. CHS and bg cells also exhibit compartmental missorting of proteins, such as elastase and cathepsin G. We have used a positional cloning approach to identify the gene that is mutated in bg mice: 1. Using intersubspecific backcross mice, bg was localized to a 0.24 cM interval on mouse Chromosome 13. 2. A physical map of 2400 kb of the bg nonrecombinant interval was generated by STS content mapping of YAC clones. 3. A candidate gene for bg, designated Lyst (LYSosomal Trafficking regulator) was identified by direct selection from a nonrecombinant interval YAC clone. 4. Lyst was disrupted by a 5kb deletion in bg<11J> mice, and Lyst mRNA was markedly reduced in bg<2J> homozygotes. Lyst encodes a novel, potentially prenylated protein with sequence similarity to stathmin, a phosphoprotein involved in intracellular transport through regulation of microtubule polymerization. The homologous human gene, LYST, is highly conserved with mouse Lyst, maps within the nonrecombinant interval for Chediak-Higashi syndrome on human Chromosome 1q42-q43, and contains a frame-shift mutation at nucleotides 117-118 of the coding domain in several unrelated CHS patients. Thus bg mice and human CHS patients have homologous disorders associated with Lyst mutations.

3. **Characterization of predicted genes from the C. elegans genome sequence by visualization of expression patterns in whole animals by FISH**

Donna Albertson
MRC Laboratory of Molecular Biology, Cambridge, United Kingdom Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Characterization of the expressed sequences predicted from the genome sequence assembled by the Caenorhabditis elegans Genome Sequencing Project is being accomplished using high resolution fluorescent in situ hybridization (FISH). In C. elegans, most somatic cells can be identified in the light microscope throughout development, allowing mRNA distributions to be visualized in individual cells in whole animals. To date post-embryonic expression patterns have been obtained for more than 90% of the approximately 200 genes tested from a sample of four contigs. Expression of the majority of the genes was seen in all cells, but only rarely were the genes expressed uniformly throughout the animal. Instead, for most of the genes, there were distinct, gene-specific patterns in the intensity of hybridization to different cells or tissues, as well as changes in expression level during development. A small proportion of genes were expressed only in specific cells or tissues. For almost all genes, transcription levels decreased with developmental age, dropping most dramatically after the molt from the fourth larval stage to adulthood. A comparison of one autosomal contig consisting of 70 genes with one from the X chromosome (65 genes) revealed little difference in gene expression patterns between the two contigs with the exception that a significantly greater number of the X-linked genes were expressed solely in muscle. Expression patterns are available in the C. elegans database, ACeDB.

4. **New ways of charting and exploring regulatory networks**

Pierre Colas
Department of Molecular Biology, Massachusetts General Hospital, Department of Genetics, Harvard Medical School, Boston, MA, USA

Fundamental cellular functions are controlled by complex networks of interacting proteins. We and others have been using interaction mating two hybrid approaches to chart these networks and determine the function of their members.

Many current approaches to the study of proteins in these networks involve overexpression or inhibition of individual protein members. We have been exploring a complementary approach, in which we use artificial proteins to disrupt the connections between these members.

Basing our design on complementarity-determining regions of immunoglobulins, we constructed a combinatorial library of constrained 20-residue peptides displayed by the active-site loop of E.coli thioredoxin. We used a two-hybrid system to isolate those that bind human cyclin-dependent kinase 2 (Cdk2). These peptide aptamers recognize different epitopes of Cdk2 with equilibrium dissociation constants in the nanomolar range. Those tested inhibit Cdk2 kinase activity by disrupting its interaction with a protein substrate. We are exploring their use in vivo to disrupt specific protein interactions.

In other experiments, we sought to construct another tool for probing protein networks that could inactivate network members and their associated proteins. To this end, we fused to the aptamers a hect domain which carries a ubiquitin-ligase activity. These fusion proteins still recognize Cdk2, albeit at much lower affinity. When co-expressed with Cdk2 in yeast, they trigger specific ubiquitylation of their target. We hope these molecules will lead to the inactivation or destruction of targeted proteins.

These results show that peptide aptamers constitute a novel kind of recognition molecule with some clear advantages above monoclonal antibodies. Peptide aptamers are designed to be expressed inside cells where they recognize specifically a protein target. They are isolated together with their coding genes. Their small size should allow them to disrupt specific protein interactions while not affecting others. Furthermore, our experiments suggest that it is possible to fuse to these "recognition modules" various "effector modules" to construct more sophisticated molecular machines. We imagine that peptide aptamers will contribute to the development of an intracellular nanotechnology aiming at destroying, modifying, moving, and assembling protein targets inside cells.

5. **The genomic route to the isolation of novel genes**

Anthony J. Brookes
Medical Research Council, Human Genetics Unit, Edinburgh, Scotland

Current large scale efforts both to sequence various model genomes, and to isolate, sequence and map random cDNAs (as ESTs) from many tissues tends to give one the impression that the task of gene identification in the human will be fully complete in a matter of months or at most a year or so. While some might argue that this is essentially true, there are other that question just how far towards a truly complete genome transcription map these efforts will really take us in advance of the human genome itself being fully sequenced. The reason behind this questioning concerns the unknown number of genes that may be missed by the EST initiative due to properties such as very low or very time or tissue restricted patterns of expression. The possibility that this number is large is strongly promoted by the observation that under one third of the 'unbiased' genes identified within the completed sequence of the s. cerevisiae genome are represented by homologues in current dbEST and other databases. In light of this there may still today be a valid role for the direct 'genomic' cloning of human genes, since this approach requires no prior assumptions about patterns or levels of expression, or of genomic location.

We have now used this approach in a search for novel genes for two gene 'families' of interest to us,

namely i) synuclein genes, and ii) genes of oxidative phosphorylation. Employing a combination of degenerate PCR and library screening approaches, both within and between species we have found that several human genes of interest to us do exist but that these are not yet represented in any public databases. Our experiences with these gene finding exercises shall be described and related to similar observations made during various focussed positional cloning exercises. The general conclusion we draw is that many genes are being missed by the EST initiative as it is currently being performed.

6. **Transcriptional map of human chromosome 21: How far from completion?**

Marie-Laure Yaspo
Max Planck Institute for Molekular Genetics, Berlin, Germany

Exon-trapping and cDNA selection techniques have been applied to chromosome 21 specific genomic clones representing about 30 megabases of random DNA. Integration of the putative new transcripts with the physical map shows a non-random gene distribution with a major clustering towards the telomere. Chromosome 21 is estimated to encode for 500-800 genes, and more than 700 independent gene fragments have been identified so far. This approach provide a way to assess the feasibility of constructing chromosome-scale transcript maps by exon-trapping and cDNA selection. If combining both techniques partly circumvent inherent biases, assembling short gene fragments into "bona-fide" discrete genes remains a challenge. The status of the actual transcript map will be discussed, together with approaches used for converting the scarce transcribed elements into genes. Reflections onto gene organization will be evoked.

7. **Evaluation of candidate tumour suppressor genes from the chromosome 1p36 region: A locus commonly deleted in several cancers**

Patrick Onyango, Barbora Lubyova, Paola Gardellin, Robert Kurzbauer and Andreas Weith
Research Institute of Molecular Pathology, Vienna, Austria

The chromosome 1p36 region recurrently reveal non random allelic deletion in a number of cancer genomes. Examples of the cancer types affected include; neuroblastoma (NB), primitive neuroecotodermal tumour (PNET), malignant melanoma, wilms tumour, ductal breast cancer, mackerel cell carcinoma, T-cell leukaemia, small cell lung carcinoma and hepatocellular carcinoma (HCC). The deleted regions are thought to contain genes essential in tumour suppression. Our focus has been to isolate the presumptive neuroblastoma, PNET and HCC suppressor genes from the 1p36 locus.

The isolation of tumour preventing elements using positional cloning approaches can be problematic when such genes are masked within large deletions. To circumvent some of these problems we first aimed at the establishment of an integral map of the region. The generation of the integral map involved creating a framework composed of a fluorescence in situ hybridization (FISH), a pulsed-field gel electrophoresis (PFGE), a cytogenetic and a genetic map consisting of YACs, P1s, PACs, and cosmids. Secondly, we determined the minimally deleted region in NB. This was achieved through FISH characterization of two small 1p36 interstitial deletions from a constitutional genotype of a NB patient, and an NB cell line. Thus, we defined a NB consensus deletion, which is roughly 3Mb in size. Incidentally, two balanced translocations found in a NB and in a PNET cell line mapped to the NB consenus deletion. We could further show that the previously proposed NB candidate suppressor genes like Id3, TNFRI, PAX7, DAN1, CDC2L all map outside the newly defined consensus deletion.

Employing exon trapping, cDNA selection, Zoo blot, and limited sequencing we have isolated 7 new genes within the NB consensus deletion and also mapped 3 previously known genes to the same locus. These genes were treated as possible candidate tumour suppressor genes for NB and PNET since they were

isolated from clones mapping in the vicinity of these two translocations. Previous loss of heterozygosity studies have also shown that HCC deletions overlap with our NB consensus deletion, and hence this tumour was included in our subsequent analysis. Full length cDNAs corresponding to these genes were isolated. Transcript characterization of the cDNAs will be presented. Briefly, sequence search of the public domain data bases using the blast software showed some interesting features for some of the genes. For example one of the genes was homologous to a tyrosine phosphatase, and another to a cell cycle checkpoint protein. To assess the possible role of these genes in the NB, PNET and HCC we analysed the integrity of these genes in the respective tumours, using FISH, southern and northern blots. A total of 18 primary NB tumours, 11 NB cell lines and 9 HCC primary tumours were analysed. None of the genes seemed to be rearranged in NB either at the genomic or at the transcript level. However, these results do not exclude the presence of subtle mutations like single base rearrangements. Analysis of the HCC tumours on the other hand revealed deletion and insertions in the cell cycle check point gene. Suggesting that the gene might have been functionally affected in the HCC tumours. The genes which showed no rearrangements in the tumours we tested may still have roles in other cancer types which frequently show rearrangements of the chromosomal 1p36 locus.

8. **Progress in the generation of an integrated physical and transcription map of human 8p11-p21 region involved in breast cancers and hematological malignancies**

Marie-Josephe Pebusque(1), Alexandra Imbert(1), Francoise Ugolini(1), Jose Adelaide(2), Max Chaffanet(1,2), Amel Dib(1), Cornel Popovici(1), Remi Houlgatte(3), Charles Auffray(3) and Daniel Birnbaum(1,2)
(1)Laboratory of Molecular Oncology, U119 INSERM, Marseille, France; (2)Laboratory of Tumor Biology, Institut Paoli-Calmettes, Marseille, France; (3)Genexpress, Villejuif, France.

The region 8p11-p21 of the short arm of human chromosome 8 is involved in several pathologies such as malignant tumors harbouring deletions, genomic amplifications or translocations, and the Werner syndrome. We have recently reported detailed physical maps of this region based on series of yeast artificial chromosomes (YACs) (Dibet al., 1995; Chaffanet et al., 1996; Imbert et al.,1996).

A high resolution transcript map is now in progress. It is obtained by a combination of several strategies which include genomic sequencing of Island Rescue PCR products, exon trapping, direct cDNA selection and mapping of chromosome 8 expressed sequence tags (ESTs). A combination of exon amplification and cDNA selection methods are applied to cosmid libraries constructed from representative YACs. Putative transcribed sequences obtained are cloned in plasmid vectors. Characterization of unique clones is performed by mapping to the corresponding genomic region, sequencing and DNA sequence analysis. For selected clones, expression studies and cDNA libraries screening are performed. In addition, expressed sequences isolated from systematic screening of human cDNA libraries (Auffray et al., 1995), and previously assigned to chromosome 8, are precisely mapped to the region of interest. The results from these studies should provide candidate genes for 8p11-p21 associated diseases.

Auffray et al. C.R. Acad. Sci. (Paris), 318: 263-272, 1995.
Dib et al., Oncogene, 10:995-1001, 1995.
Chaffanet et al., Cytogenet. Cell Genet. 72:63-68, 1995.
Imbert et al., Genomics, 32: 29-38, 1996.

9. **2,006 ESTs from human chromosome 7-enriched direct cDNA selection libraries**

Jeffrey W. Touchman(1), Gerard G. Bouffard(1), Luping Wang(2), Lauren A. Weintraub(1), Jacquelyn R. Idol(1), Jesse C. Nussbaum(1), Christiane M. Robbins(1), Michael Lovett(2), and Eric D. Green(1)

(1)National Center for Human Genome Research, National Institutes of Health, USA; (2)The University of Texas Southwestern Medical Center, USA

The establishment and mapping of gene-specific DNA sequences are highly complementary to the ongoing efforts to map and sequence all human chromosomes. To facilitate our studies of human chromosome 7, we have generated and analyzed 2,006 expressed-sequence tags (ESTs) derived from a collection of direct selected cDNA libraries highly enriched for human chromosome 7 genes. Similarity searches indicate that approximately two-thirds of the ESTs are not represented by sequences in the public databases, including those in dbEST, and thus represent new gene sequences. In addition, a large fraction do not have redundant or overlapping sequences within our collection. PCR assays (i.e., STSs) have been developed for 190 of these ESTs. Remarkably, 181 out of 190 (96%) of these STSs mapped to chromosome 7, demonstrating the robustness of the chromosome enrichment in the construction of the direct cDNA selection libraries. Thus far, 140 of these EST-specific STSs have been unequivocally assigned to YAC contigs from different parts of the chromosome. Together, these studies provide over 2,000 ESTs highly enriched for chromosome 7 gene sequences, 181 new chromosome 7 STSs corresponding to ESTs, and a definitive demonstration of the ability to enrich for chromosome-specific gene sequences by direct cDNA selection. Furthermore, the libraries, sequence data, and mapping information generated to date should greatly enhance construction of the chromosome 7 transcript map.

10. **Large-scale functional analysis of the S. cerevisiae genome**

Petra Ross-Macdonald
Yale University, New Haven, CT, USA

Now that the genome of S. cerevisiae has been sequenced, the focus of research has moved towards investigating the function of the encoded proteins. In a pilot project, we used transposon mutagenesis to create a bank of yeast strains, each with a lacZ insertion at a random genomic location. The bank is the basis of a 3-in-1 approach for determining when genes are expressed during the life cycle, the subcellular locations of their encoded proteins, and the phenotypic effect of disrupting the gene (Genes Dev. 8:1087-1105, 1994). We identified 3,600 strains with fusions expressed during vegetative growth, mating or meiosis. The fusion protein localizes to a discrete site in 12% of these strains. Phenotypes were analysed for 186 insertions, and the disrupted gene identified in 263 strains. In an extension of this work, we have used preparations of spread nuclei to examine strains with lacZ-fusion proteins that localize to the nucleus. The fusion protein can be immunolocalized to discrete sites on chromosomes in 20 of 31 strains examined. Of 9 strains carrying fusions in known genes, 5 are in known transcription factors and DNA binding proteins.

We have now initiated a full-scale effort to analyze all ORFs in the yeast genome by this approach. This project utilizes a new library and novel adaptations of large-scale techniques pioneered in genome projects. To improve our ability to immunolocalize the products of the mutagenized genes, the transposon used (mTn-3xHA/lacZ) has been extensively modified. mTn-3xHA/lacZ contains the lacZ gene, allowing us to identify in-frame fusions as before. However, a lox site is present near each end of the transposon. Upon expression of the Cre recombinase, all sequence elements between the lox sites are excised, leaving a single lox site and flanking transposon sequences. We have engineered these flanking sequences to contain three tandem copies of the 'HA' epitope from influenza virus hemagglutinin. Thus the excision event generates an in-frame insertion of 93 amino acids (a 'HAT' tag) in the product of the mutagenized gene.

mTn-3xHA/lacZ has performed very well in test-mutagenesis of two yeast genes. The HAT-tagged proteins were functional and localized correctly with the majority of insertions. We are further modifying the transposon to incorporate selection markers allowing its use in other eukaryote systems, and have also

created a version allowing generation of in-frame fusions to the green fluorescent protein.

In summary, our approach will provide insight into the function of both novel and previously characterized yeast gene products. The results of our new project will be deposited in a publicly accessible data base (http://ycmi.med.yale.edu/YGAC/home.html), allowing researchers who identify a yeast gene (or a homolog thereof) to know when that gene is expressed, whether its gene product localizes to a specific subcellular location, and the phenotype of the insertion mutant.

SESSION 5

1. **Novel characteristics of the 3q21 leukemia breakpoint region**

A. Rynditch(1,2), Y. Pekarsky(2), and K. Gardiner(2)
(1)National Academy of Sciences of Ukraine, Kiev, Ukraine (2)Eleanor Roosevelt Institute, Denver, CO, USA

Human chromosome band 3q21 is frequently involved in rearrangements associated with leukemia. Rearrangements include deletion of 3q21, inversion, insertion and translocation of 3q21:3q26, and translocation of 3q21 to 1p35, 9q22 and 11q23. Cloning and characterization of the major 3q21 breakpoint region has revealed several novel features:

i) high novel gene density: Exon trapping, cDNA selection and genomic sequencing followed by Grail analysis have so far identified fragments of 10 genes within an approximately 100 kb segment. This is the highest demonstrable gene density associated with any leukemia breakpoint region. With the exception of the Ribophorin I gene, all cDNAs are novel.

ii) low transcriptional activity: With one exception, the novel genes are transcribed at relatively low levels. Transcription of 5 genes is detectable only by RT-PCR; only 1 of 5 tested is present in a cDNA library.

iii) dispersion of breakpoints: Positions of 10 leukemia breakpoints, including t(3;3), inv3 and (1;3), have been correlated with the transcriptional map. Breakpoints are distributed throughout a 35 kb segment spanning 3 genes, suggesting that, contrary to other AML cases, it is unlikely that a single fusion protein can be involved.

iv) altered transcriptional activity: Preliminary data from a leukemia-derived cell line carrying a t(3;3) indicates increased expression of several genes mapping in the vicinity of the breakpoint. It is postulated that the rearrangement produces a generalized deregulation or destabilization of transcription.

2. **Drosophila-related expressed sequences (DRES): A source of candidate genes for human diseases**

G. Borsani(1), S. Banfi(1), E. Rossi(2,3), A. Bulfone(1), L. Bernard(1), F. Rubboli(1), A. Marchitiello(1), A. Guffanti(1), G. Simon(1), S. Giglio(3), E. Coluccia(3), M. Zollo(1), O. Zuffardi(2,3), A. Ballabio(1,4)
(1)Telethon Institute of Genetics and Medicine (TIGEM) (2)Servizio di Citogenetica, San Raffaele Biomedical Science Park, Milan, Italy; (3)Cattedra di Biologia Generale, University of Pavia, Pavia, Italy; (4)Department of Molecular Biology, University of Siena, Siena, Italy.

Cross-species comparison is an effective tool used to identify genes and study their function in both normal and pathological conditions. We have applied the power of Drosophila genetics to the vast resource of human cDNAs represented in the EST database (dbEST) to identify novel human genes of high biological interest. Sixty-six human cDNAs showing significant homology to Drosophila mutant genes were identified by screening dbEST using the text string option, and their map position was determined using both fluorescence in situ hybridization (FISH) and radiation hybrid mapping. These cDNAs, which were termed DRES (Drosophila-related expressed sequences), represent positional candidate genes for human diseases mapping to the corresponding genomic region. For example, DRES9 is homologous to the Drosophila retinal degeneration B gene and was mapped to 11q13.5, where at least three types of human retinopathies were assigned. We are now focusing on identification and mapping of additional DRES, isolation and sequencing of full-length transcripts, identification of yeast artificial chromosomes (YACs),

isolation of murine homologs of DRES, genetic mapping in the mouse, and expression studies in mouse embryo by RNA in situ hybridization experiments. Comparison between DRES genes and their putative partners in Drosophila and mouse may provide important insights into their function in mammals and their possible role in human disease.

3. En masse terminal exon trapping of the human Y chromosome

Yun-Fai Chris Lau
University of California-San Francisco, Division of Cell and Developmental Genetics, San Francisco, CA, USA

The identification of ESTs from the human Y chromosome is a difficult task because most genes on this chromosome are postulated to be either specific for male physiology and development or members of polygenic traits whose expression are unknown. So far, random sequencing efforts have not been fruitful in generating ESTs from this chromosome. As a first step to collect ESTs from the human Y chromosome for the construction of a transcription map, we have applied the 3 terminal exon trapping technique to identify ESTs en masse from this chromosome. Two approaches have been adopted for this study. First, total DNA derived from an entire Y chromosome cosmid library (with >13,000 cosmids) was digested with a trapping restriction enzyme and ligated to a trapping cassette derived from the vector, pTAG4. The ligated DNA was then transfected transiently to COS7 cells and the resulting transcripts were amplified with 3 RACE and primers specific for the 3 terminal exon trapping vector. Second, a novel technique has been developed to transfer en masse the Y chromosome DNA from the initial cosmid library (in Lawrist 16) to another cosmid trapping vector, pTAG5. This procedure inserts the Y chromosome DNA immediately downstream of the 3 terminal exon trapping cassette, thereby the transcription of composite transcripts suitable for the exon trapping procedure. Analysis of the resulting exons by PCR using specific primers of 9 genes on this chromosome indicated that all were present in the final exon products which were subsequently subcloned into pAMP1 plasmid and arrayed into 51 96-well microtiter dishes. Hybridization analysis indicates that 8.7% and 2.8% of these exon clones were derived from two repeated genes, TSPY and RBM, on the short and long arms respectively of this chromosome. Random sequencing of 170 individual clones confirmed the frequencies of both genes. Further, over 54% of the clones were either homologous to sequences of genes located on this chromosome or unique sequences having characteristics of valid 3 terminal exons. Sequence analysis of 14 TSPY exon clones indicated that 57% of them harbor 2-4 internal exons in addition to the terminal exon. The results indicate that: 1) en masse 3 terminal exon trapping for an entire chromosome is highly feasible, 2) >50% of resulting exon clones are either derived from known Y genes or potential functional sequences from this chromosome, and 3) similar strategy should be applicable for en masse isolation of ESTs from other human chromosomes.

4. Transcript mapping in the nonrecombinant interval for the mouse neuromuscular disease gene mnd2 and the corresponding region of human chromosome 2p13

Miriam Meisler, Wonhee Jang and John S. Weber
Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA

The mnd2 mutation in the mouse causes a lethal muscle wasting disease. Affected mice develop an unsteady gait with uncontrolled muscle contraction and die by 40 days of age. The mnd2 gene is located within a conserved linkage group on mouse chromosome 6 that corresponds to human chromosome band 2p13. In order to isolate this disease gene, we generated a high resolution genetic cross that localized the gene to an interval of 0.2 +/- 0.1 cM, which in the mouse genome is equivalent to approximately 400 kb. Within and around the nonrecombinant interval, we have mapped nine positional candidate genes from human chromosome 2p13 and four novel genes isolated by exon trapping of P1 clones. The complete

sequences of the coding regions of six genes within the nonrecombinant region, from homozygous mutant DNA and the strain of origin, has not detected any mutations. The results of Northern blotting indicate that all of these transcripts are present at normal levels in mutant tissues.

The novel genes identified in the region include one new member of the WD-repeat family and one evolutionarily conserved gene with 35% sequence identity to a gene in C. elegans. To progress from individual exons to full length cDNAs we have relied heavily on 5' and 3' RACE and vector-insert PCR from cDNA libraries. For some genes, identification of related human ESTs has been an important tool. Our results indicate that this is a gene rich region with conserved gene order in human and mouse that would be a good substrate for comparative sequencing. A comparative gene map of the human and mouse chromosome regions will be presented.

5. **New genes within the MHC: Cloning by paralogy: The case of the B30-gene family**

Ribouchon M.T., Offer C., Matte M.G., Tazi Hanini R., Henry J., and Pontarotti P.
INSERM U119, Marseille, France

We describe an easy method to identify new genes within a given genomic region. This concept is based on the fact that multigenic family members are derived from a common ancestor and that some duplicate genes stay syntenic (linked) even after several hundred of million years. Our pilot experiment was performed on the RFB30 multigenic family. During the search of new coding sequences within chromosomal region which contains the human histocompatibility complex, we have found an exon encoding a protein domain (170 amino acids) named B30.2. This domain was found associated with 3 different N terminal domains in few other proteins 1) A Ring finger domain as in the case of RFP 2) An IgV-IgC like domains in the case of butyrophilin (BT) 3) A leucine Zipper domain. Several of these proteins were colocalised in the same Chromosomal region (Vernet et al., 1993, Amadou et al., 1995). The accumulation of new information in dbEST suggested us the possibility to use it for the screening to obtain the cDNA of other members of this large family. We have screened the dbESTdata base using the BLASTx algorithmic software with the amino-acid sequence of the RFP and BT protein. 31 ESTs showed a similarity with the RFP and BT proteins. Some clones showed overlapping sequence meaning that they come from the same transcription unit . The clones for which we did not found similarities or overlapping sequence with other selected ESTs were used to hybridize a Southern blot containing genomic DNA digested by different restriction enzymes. The same pattern of hybridization found with two different ESTs means that they come from the same genomic region and thus correspond probably to the same gene. After the sequencing of the different ESTs, transcripts consensus (TC) were constructed corresponding eventually to full length transcript. Eight TCs were selected by this methodology. Chromosomal localization was performed by in situ hybridization on chromosome. Three of the TCs were localized to the chromosome 6 in the 6p21.3-6p22 band, thus likely close or within the MHC. They were more precisely map using Yeast Artificial Chromosome (YAC) covering the MHC and its telomeric part. Therefore the cloning by paralogy could be an efficient approach to clone new genes within a given chromosome region.

Vernet C.et al., J Mol Evol 37:600-612, 1993 Amadou C. et al., Genomics, 26:9-20, 1996

6. **The isolation of CpG islands from human chromosomal regions 11q13 and Xp22 on the basis of thermal stability of DNA fragments covering these regions**

Masahiko Shiraishi
Oncogene Division, National Cancer Center Research Institute, Tokyo, Japan

We have developed a method for the preferential isolation of DNA fragments associated with CpG islands

on the basis of thermal stability of the fragments (segregation of partly melted molecules, SPM). SPM has been applied to isolate CpG islands from the human chromosomal regions 11q13 and Xp22. The former is a CpG-rich region (R band), with the latter being less CpG-rich (G band). Nine P1 clones covering a 500-kilobase (kb) region of 11q13 and ten cosmid clones covering a 300-kb region of Xp22 were digested with four restriction endonucleases, MseI, BfaI, NlaIII, and Tsp509I. These restriction enzyme digests were then subjected to denaturing gradient gel electrophoresis. About sixty fragments from 11q13 P1 clones and five fragments from Xp22 cosmid clones were obtained as retained ones. Nucleotide sequence analysis revealed that many of the recovered fragments from the 11q13-derived clones contained CpG islands, including that of the ADRBK1 gene, which has previously been assigned to this region. However, none of five from Xp22 clones possessed properties for CpG islands. These results suggest that SPM facilitates the efficient isolation of CpG islands reflecting CpG density of the corresponding region.

7. **A gene rich region around the ATM locus - Physical map and identification of two new transcripts in the 11q23.1 region** N. Udar(1), X. Chen(1), S. Xu(1), J. O. Bay(1), S. Dandekar(1), T. Liang(1), N. Uhrhammer(1), H. Shizuya(2), H. Yang(1), P. Concannon(3), L. Yang(1), Z. Wang(1), M. Telatar(1), R. A. Gatti(1)
. (1)Department of Pathology, UCLA School of Medicine, Los Angeles, CA, USA (2)Division of Biology, California Institute of Technology, Pasadena, CA, USA (3)Virginia Mason Research Center, Seattle, WA, USA

The gene for the autosomal recessive neurological disease - Ataxia Telangiectasia (ATM) has recently been identified. An international consortium of eight laboratories localized the A-T gene by linkage analysis of 176 families to chromosome 11q23.1, between the markers D11S384, and D11S535. We constructed a series of contigs using three BACs and twelve cosmids, spanning this region of approximately 400 kb. We also developed twenty-one STS markers from the BACs and cosmids. With all this information we have been able to build a precise map of the contig with respect to the exons for ATM and STSs. This information will be useful for further studies of functional domains and regulatory functions within the ATM gene.

Two new transcripts were isolated by cDNA selection techniques, CAND3 and NS6. Both these transcripts are immediately proximal to the ATM gene. CAND3 spans ~140 kb of genomic DNA, and is located immediately centromeric to ATM, with 544 bp separating the two genes. The two ubiquitously-expressed transcripts, ATM and CAND3, are transcribed in opposite directions, sharing a common promoter region. The predicted protein has weak homologies to transcriptional factors, nucleoporin protein, and protein kinases. No homology to ATM, nor any mutation of CAND3 in A-T patients, have been found. The 5 to 5 orientation of CAND3 and ATM suggests co-regulation of biologically-related functions. The other transcript, NS6, is a novel sequence with near perfect homology with the catalytic domain of the sodium glucose cotransporter gene family. This gene localizes proximal to CAND3 and the MAT (ACAT) gene. This makes a total of 4 genes within a 1 Mb region.

8. **Analysis of differentially expressed genes using solid-phase approaches in transcript selection and in characterization of cognate proteins expressed in bacterial systems**

Jacob Odeberg, Magnus Larsson, Mathias Uhl�n, Stefan Sthl and Joakim Lundeberg.
Department of Biochemistry, KTH, Royal Institute of Technology, S-100 44 Stockholm, Sweden, Phone: (int)-46-8 790 87 58; Fax: (int)-47-8 24 54 52.

Solid-phase methods based on both differential display (DD) and representational differential analysis (RDA) have been designed enabling differential gene expression analysis in samples with scarce amounts of mRNA originating from stem cells, sperm cells and colon tissue. The principle has been to utilize the

streptavidin biotin system to capture nucleic acids onto magnetic microbeads and combine that with in vitro amplification techniques. The immobilization of mRNA and/or DNA fragments to a solid phase has significantly simplified the purification procedure and minimized sample loss that may also facilitate future automation. To investigate the biological function of the isolated fragment a bacterial expression approach has been designed for the production of cognate gene fragments. Specific bacterial expression vectors has been constructed to create fusion proteins between isolated fragments and affinity protein tags of protein A and protein G origin. Fusion proteins, generated after E. coli propagation and affinity chromatography purification, have been employed in the subsequent eliciting of specific antibodies enabling various functional analysis in the monitoring of expressed genes.

9. **Continuing studies on the subcellular localization of mRNAs in neurons**

Jennifer Phillips, Kevin Miyashiro, Peter Crino and Jim Eberwine
University of Pennsylvania Medical School, Philadelphia, PA, USA

In both the developing and mature neuron, plasticity of the cell is in part determined by the activities of neuronal processes. Neuronal processes (axons and dendrites) establish the polarity of the neuron and subserve specialized functions, such that dendrites provide information to presynaptic cells and axons provide information to post-synaptic cells. The factors which account for the specialized functions of neuronal processes has been an area of intense research and has lead to the characterization of certain proteins which are selectively enriched in these structures. Additionally, in situ hybridization has localized several mRNAs to dendrites and a few mRNAs in axons of specific neuronal subtypes. Recently, we have used more sensitive techniques (single cell nucleic acid amplification, expression profiling, and sequencing of cDNAs derived from differential display) to analyze the mRNAs present in processes and have greatly expanded the known repertoire of mRNAs. We have also been developing a more sensitive technique to determine proteins in individual processes (immuno-aRNA). This approach will provide information on the mRNA/protein profile of individual processes and their developmental regulation and potentially insights into the mechanisms of plasticity.

10. **Successful modeling of the functional organization of transcriptional units**

Thomas Werner
Forschungszentrum fuer Umwelt und Gesundheit GMBH Institut fuer Saeugetiergenetik,,
Oberschleissheim, Germany

Regulatory regions of higher eukaryotes usually encompass multiple regulatory elements which exert their regulatory function only within the correct context. Similarity of regulatory regions often is not evident on sequence level precluding global alignment strategies for their detection. We used a strictly modular concept to deduce the functional organization of regulatory regions from a set of such regions. This concept - implemented in the program Model Generator - allows development of highly specific models from a training set of 10 to 20 sequences. Only a simple initial model (e.g. two characteristic transcription factor binding sites) is required. Construction of the model can reveal new common elements. Therefore, model development also allows delineation of new potential functional elements for targeted experimental verification in addition to the generated model. Another program designed ModelInspector has been developed to scan new sequences for matches to these models.

We present two models for retroviral transcriptional control regions which contain functional pol II promoters. Both models (C-type LTR and Lentivirus LTR, e.g. HIV) were used with the program ModelInspector in order to assess the specificity of the models. All LTRs from both training sets and 18 LTRs of other types (B- type, D-type, Spuma, HTLV I/II) were matched against the models. Both models

fully recognized their respective training set and clearly discriminated against all other LTR types. The program ModelInspector is also able to carry out database searches with these models. Tests with the C-type LTR model have shown that ModelInspector did not miss a single known LTR (= no false negatives). The program revealed 5 new C-type LTR candidates in addition to 17 known full length LTRs in the primate section of GenBank Release 92.0 (about 95 million nucleotides on both strands). The first of these candidates has been experimentally verified as a weak (pol II) promoter. This agrees with the low scoring of this LTR which also suggested weak activity and demonstrates the predictive power of organization-based models of transcriptional units in mammalian genomes.

Poster Abstracts

1. **Identification and characterization of human and murine homologs of Drosophila eye mutant genes**

S. Banfi(1), G. Borsani(1), A. Bulfone(1), L. Bernard(1), F. Rubboli(1), A. Marchitiello(1), M. Zollo(1), O. Zuffardi(2,3), A. Ballabio(1,4).
1Telethon Institute of Genetics and Medicine (TIGEM) (2)Servizio di Citogenetica, San Raffaele Biomedical Science Park, Milan; (3)Cattedra di Biologia Generale, University of Pavia, Pavia; (4)Department of Molecular Biology, University of Siena, Siena, Italy

In spite of the considerable evolutionary distance, the mechanisms of eye development and vision present many common features in vertebrates and invertebrates. A large number of human and mouse genes concerned with the function or development of the eye have been shown to be highly conserved throughout evolution and to have closely similar orthologs in Drosophila melanogaster. In the course of a project aimed at the recognition of human cDNAs homologous to Drosophila mutant genes through EST database (dbEST) searching, we identified twelve human cDNAs homologous to Drosophila eye mutant genes. The degree of homology shared by these DRES (Drosophila-related expressed sequences) genes and the corresponding Drosophila eye mutant genes is very significant with a Pvalue at the protein level <10-13 in all cases.

The isolation and sequencing of the full-length human and murine transcript of these DRES is currently in progress. Furthermore, the expression pattern of DRES genes is being studied by RNA in situ hybridization in mouse embryo standard sections. The analysis of the data will provide useful information on the putative function of these genes in vertebrates and their possible involvement in human eye inherited disorders. Moreover, the correlation with the expression pattern of the corresponding Drosophila genes will be helpful to assess a conserved function of these genes during evolution.

2. **Identification of genes involved in haemopoietic stem cell self-renewal and commitment**

Sarah Baxendale, Lynne Hampson, Ian N. Hampson and T. Michael Dexter
Paterson Institute of Cancer Research, Department of Experimental Haematology, Christie Hospital, Manchester, United Kingdom

Haemopoietic stem cells have the ability to either self-renew or to differentiate into all the mature blood cell types. The mechanisms underlying these processes are, as yet, unclear but they are ultimately driven by changes in gene expression. Using the chemical cross linking subtraction technique CCLS (see also abstracts by J.M. Walter et al and improvements to CCLS by Hampson et al) we have now isolated a cohort of genes which down-regulate during myeloid differentiation of the well characterized, non-leukaemic murine multipotent stem cell line FDCP-Mix. Sequence analysis of these clones continues to identify both known and novel cDNAs. Future functional studies will allow determination of the precise role which the protein products of these genes have in haemopoietic stem cell self renewal and differentiation and should further our understanding of normal haemopoiesis and as a corollary, leukaemogenesis.

3. **Transcript mapping in a gene-poor region of the human X-chromosome**

Brooksbank, R.A., Coffey, A.J., Cahn, A.P., Rhodes, S., Vaudin, M., Howell, G.R, Micklem, G., King, A., Durham, J., Bye, J.M., Dunham, A., Ross, M.T. and Bentley, D.R.
The Sanger Centre, Wellcome Trust Genome Campus, Cambridge, United Kingdom

Now that the Human Genome Project has entered the large-scale sequencing stage, the identification of all the genes is a major goal. We are studying an approximately 2.5 Mb region of Xq25 delineated by a deletion present in a male patient suffering from X-linked lymphoproliferative syndrome (XLP), the gene responsible for which has not yet been identified. This region has been extensively mapped and a bacterial clone contig is being assembled across it. A minimal overlapping set of clones from this contig are being sequenced. The absence of any other symptoms in the patient implies that the 2.5 Mb region contains at most only a few genes, making it a potentially challenging region for transcript mapping. We are using exon-trapping, cDNA selection and direct screening of cDNA libraries with genomic clones to identify putative genes within the region. We have used an acedb database to collate the results of these experiments along with exon predictions and homologies to known sequences. The genomic sequence is then used to design PCR primers to screen cDNA libraries; positive clones are themselves sequenced and incorporated into the database. The abundance of repetitive elements in the region is a recurrent obstacle to all the methods that we are using. We can use the sequence data to visualize the distribution of repeats and to interpret our results.

4. **Physical mapping and identification of expressed transcripts within the region of 13Q14.3 commonly deleted in B-cell chronic lymphocytic leukemia**

Martin M. Corcoran, Roman Mullenbach, Rachel E. Ibbotson, Anne Gardiner and David G. Oscier
Molecular Biology Laboratory, Department of Pathology, Royal Bournemouth General Hospital, Bournemouth, United Kingdom

A region of human chromosome 13Q14.3 commonly deleted in patients with Chronic Lymphocytic Leukemia has been subcloned into a cosmid and pac contig. The area covered, over 600 kb in total, including the markers D13s25 and D13s319 has been physically mapped and the positions of several CpG islands determined, possibly associated with a putative tumor suppressor gene at this locus. The basis for a transcriptional map of the region has been constructed by screening multiple tissue northern sources with cosmid inserts and a number of expressed transcripts have been identified. The cosmid and pac contig has been used as a source of microsatellite markers in order to identify informative polymorphisms and as a source of FISH probes leading to a narrowing down of the minimally deleted region in patient samples as an important step towards the identification and isolation of the candidate tumor suppressor gene or genes.

5. **Gene identification in polygenic disease**

Roger D. Cox
Wellcome Trust Centre for Human Genetics, Oxfordshire, United Kingdom

My group is directly involved in gene identification on a number of polygenic disease projects within the Wellcome Trust Centre for Human Genetics, Oxford. In addition part of my group is working, within an international collaboration, on the cloning of maturity onset diabetes of the young 3 (MODY3) gene- a dominant disease gene.

We are using exon trapping, cDNA selection, direct screening of cDNA libraries with genomic DNA and more recently direct sequencing.

Our experience of these approaches and of characterizing transcripts identified will be discussed.

6. **Transcriptional mapping of chromosome 16p12.3-12.2**

Treasa A. Creavin(1), Jaimie A, Greenham(1), Norman A. Doggett(2), Sara E. Mole(1)

(1)Department of Paediatrics, The Rayne Institute, University College Medical School, London, United Kingdom (2)Life Science Division and Centre for Human Genome Studies, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

As part of a European consortium contributing to the physical and transcriptional mapping of the short arm of chromosome 16, we are constructing a transcriptional map of 16p12.3-12.2. At present this region consists of 68 cosmids which map to 5 hybrid intervals. Initially our strategy involved isolating unique restriction fragments from cosmids that do not hybridize to human placental DNA.

Each cosmid was analysed for the presence of a CpG island by digestion with NotI, AscI, EagI and BssHII. We have used these results to select cosmids which we have used in Direct cDNA Selection experiments. We are currently analysing over a thousand selected cDNA clones by sequencing and hybridization with chromosome 16 clones. To date seven cDNA clones have been shown to have high homology to ESTs in the database, one cDNA clone has high homology with a gene not previously mapped to chromosome 16 and another cDNA had no homology with any genes or ESTs currently deposited in the databases.

7. **Strain-specific expression of selenium binding protein genes (Lpsb) in mice**

Laura de Gregorio, Manuela Gariboldi, Giacomo Manenti, F. Stefania Falvella, Manuela Rota, Marco A. Pierotti, and Tommaso A. Dragani
Instituto Nazionale Tumori, Milan, Italy

Susceptibility to hepatocarcinogenesis in mice represents a genetic model of polygenic inheritance of predisposition to cancer, because six Hepatocarcinogen susceptibility (Hcs) and two resistance (Hcr) loci have been mapped. With the aim of testing possible strain-specific pattern of gene expression that could be associated with genetic predisposition to liver cancer, we have analysed by mRNA differential display the normal liver tissue of different mouse strains. Two selenium binding liver protein genes (Lpsb1 and Lpsb2) were expressed in the normal liver of adult C57BL/6J and BALB/c mice (genetically resistant to hepatocarcinogenesis) at >10 fold higher levels than in the normal liver of C3H/He and CBA/J strains (susceptible to hepatocarcinogenesis). In F1 hybrid mice obtained from C3H/He strain crossed either with C57BL/6J or BALB/c mice, expression levels of Lpsb genes correlated with recessivity and dominance of resistance to hepatocarcinogenesis. The Lpsb genes are closely linked and map on Chromosome 3, in a region where no Hcs or Hcr loci have been localized.

8. **Transcriptional mapping of human chromosome 16p12.1**

Jaimie Greenham
University College, London Medical School, The Rayne Institute London, United Kingdom

As part of a European Consortium contributing to a physical and transcriptional map of the short arm of chromosome 16, we are constructing a transcriptional map of human chromosome 16p12.1. This region at present consists of 136 cosmids, representing about 50-60% coverage of the region. Our initial strategy involved identifying cosmids that contain rare-cutter CpG-rich restriction sites and isolating unique restriction fragments that do not hybridize to human placental DNA. These potential coding fragments have been used to screen multiple tissue northern blots and cDNA libraries to identify transcribed sequences. More recently direct cDNA selection and exon trapping have been carried out on a subset of these cosmids. Using a combination of these approaches we have mapped two known genes (PRKCB1 and PRKCB2), two novel cDNAs and three novel exons to the region.

In addition we have integrated other known genes (SCNN1B and SCNN1G) and IMAGE cDNAs that had

previously only been localized to somatic cell hybrid intervals spanning the region. We have also identified a Fugu rubripes cosmid that we believe to contain PRKCB1, we aim to confirm this by shotgun sequencing and in the process identify other genes that may map within close proximity.

9. **Rapid identification of fragments of human transcripts from a defined chromosomal region: Representational Difference Analysis of somatic cell hybrids**

Peter C. Groot and Bernard A. van Oost
Department of Clinical Sciences of Companion Animals, University of Utrecht, Utrecht, The Netherlands

We have used Representational Difference Analysis on cDNA's (cDNA-RDA, Hubank and Schatz, Nucl. Acids Res. 22, 5640-8, 1994), as a means to directly isolate expressed sequences derived from human Xq28. To this end, the hamster cell-line Y21 was used as Driver and the Y21-derived hamster-human hybrid Q1Z cell-line (Warren et al., PNAS 87, 3856-60) as Tester. From both cell-lines cDNA was made, digested with the four-cutter DpnII (GATC) to obtain maximum representation, ligated to linkers and amplified by PCR. Two rounds of subtraction were used with Driver:Tester ratio's of 1:100 and 1:800, in the first and second round, respectively. To monitor the subtraction procedure, we used hybridization with a fragment of the human Xq28 derived QM gene, which was known to be highly expressed in Q1Z, as probe. After one round of subtraction we observed a smear of DNA after EtBr staining, with a number of discrete bands already clearly visible, indicating that the complexity of the DNA was already substantially reduced. After two rounds of subtraction only two major bands, with a very faint background, were seen after EtBr-staining, one of which was derived from the QM-gene. This result shows that the subtraction procedure in our hands has been surprisingly effective. To investigate the nature of the second major band and to establish if more human-derived cDNA's can be isolated using this procedure, we are currently characterizing the first and second round subtraction products by subcloning and sequencing.

10. **Motif capture: Isolation and characterization of large cDNA families with GeneTrapper, zinc fingers, CAG repeats, SH2 domains, RNP1 domains**

Christian Gruber(1), Wu-Bo Li(1), Joel A. Jessee(1), Kevin G. Becker(2), Wayne D.L. Bentham(2), James W. Nagle(2), Ameer Gado(2), and William E. Biddison(2) (1)Life Technologies, Inc., Gaithersburg, MD, USA; (2)Neuroimmunology Branch and Laboratory of Neurogenetics, National Institute of neurological Diseases and Stroke, National Institutes of Health, Bethesda, MD, USA

A human brain cDNA expression library was greatly enriched for specific gene families with the GeneTrapper cDNA positive selection system coupled to motif specific degenerate oligonucleotides. Motif capture uses degenerate oligo-nucleotides coding for short highly conserved regions of gene families. Enriched sublibraries were isolated for C2H2-type zinc fingers, ribonucleoproteins (RNP1) SH2 domains, and CAG triplet repeats. These clones were filter arrayed creating family specific sets. Subsets of each family were EST sequenced to confirm the efficiency of cDNA family isolation and for identification of novel genes. Family specific filter sets are being analyzed for relative gene expression by tissue specific differential hybridization. These enriched plasmid sublibraries are being used in expression analyses linking large scale cDNA isolation with biological function. Similar family specific isolations are being performed in the yeast two-hybrid system and other specialty cDNA libraries. The motif capture approach describes a focused, general, rapid, non-random approach to genome transcript analysis. It is non-PCR based, results in large or full length cDNA clones, is easily linked to specific biological assays, and is rapidly adaptable to non-human systems.

11. **Automating the TOGA method**

K.W. Hasel(1), B. Hilbush(1), D. Seto(1), P. Miller(1), and J.G. Sutcliffe(2)
(1)Digital Gene Technologies, La Jolla, CA, USA (2)Dept of Molecular Biology, The Scripps Research Institute, La Jolla, CA, USA

The TOGA (Total Gene expression Analysis) method when practiced manually is robust but cumbersome due to the large number of PCR reactions to be performed and analyzed so as to achieve complete coverage of gene space. The method has been optimized for each of the 256 TOGA primers. We have automated the method to increase throughput and quality control, and to facilitate automatic data collection. A series of robot-controlled steps has allowed consistent and rapid generation of TOGA data, as will be illustrated with experimental examples. The unique digital address for each mRNA, a combination of an 8-nucleotide sequence and fragment length, renders the data amenable for comparison and archiving into a relational database for query. This feature is easily merged with public genome databases, allowing the convergence of gene expression data, nucleic acid sequence data, gene mapping data and literature citations.

12. Towards positional cloning of the mouse high growth (hg) gene

Simon Horvat, Roslin Institute, Edinburgh, Scotland, and Jaun F. Medrano, Dept. of Animal Science, University of California, Davis, Davis, CA 95616.

The high growth (hg) gene in mice produces a 30-50% increase in weight gain of homozygous individuals. Previous interval mapping and test crosses analyses positioned hg to ~ 5 cM interval around marker D10Mit41 on mouse chromosome 10 and demonstrated that hg is not an allele of Insulin-like growth factor I (Igf-1) or Decorin (Dcn), two closely linked candidates.

Recently, a microsatellite marker D10Mit69 was found to be deleted in high growth mice. This marker cosegregated with hg in a cross of congenic strains C57BL/6J-hghg X C57BL/6J (1051 F2 mice) indicating that a deletion of a region around D10Mit69 might be responsible for the high growth effect. Marker D10Mit69 has been utilized as an entry point to physical cloning of the hg-containing segment using Yeast Artificial Chromosome (YAC) and Bacterial Artificial Chromosome (BAC) clones. Exon trapping of the BACs is being used to uncover expressed sequences. The current status of candidate transcription units isolated from the hg region will be presented.

13. CpG islands are concentrated on chicken microchromosomes

Heather A. McQueen, Giorgia Siriaco, Sally H. Cross, and Adrian P. Bird
University of Edinburgh, Institute of Cell and Molecular Biology, Edinburgh, Scotland

We are engaged in the study of CpG islands and their organization in the chicken genome. CpG islands are short unmethylated CpG-rich sequences situated to the 5' end of many vertebrate genes and are particularly GC-rich in the chicken. The chicken karyotype comprises 39 chromosome pairs of which at least 29 are microchromosomes. Microchromosomes account for only about 25% of the genomic DNA and are cytologically indistinguishable. Due to technical limitation microchromosomal genes are poorly represented in current genome maps. However, we have used a CpG island library, prepared from chicken, to demonstrate that CpG islands are concentrated on the microchromosomes. Our results imply a higher gene density for chicken microchromosomes than is seen in mammalian species. We are currently testing this hypothesis using cloned segments of microchromosomal chicken DNA.

14. Exon-trapping in a region of the mouse Y chromosome critical for spermatogenesis

Michael Mitchell
INSERM, Marseille, France

The Sxrb deletion interval of the mouse Y chromosome short arm is critical for normal spermatogenesis and expression of the male specific minor transplantation antigen H-Y. Mice deleted for this region manifest a block in spermatogenesis 3 days after birth, during the mitotic proliferation of the germ cells. We have established contigs of cosmids and yeast and bacterial artificial chromosomes covering over 1 Mb in this region. With the aim of defining a comprehensive transcriptional map of this interval, 400 kb of these contigs have so far been subjected to exon trapping and a preliminary exon map established. Exon-trapping has been carried out using the standard pSPL3 protocol on individual cosmids. Potential exons have been designated from among 30 distinct exon-trapped products on the basis of the following criteria: 1) the sequence shows no homology to a known repeat element, 2) the sequence does not hit several unrelated mouse sequences 3) one of the forward reading frames is open 4) the sequence is either homologous to the coding region of a known gene, or shows no identity to rodent sequences, 5) PCR from genomic DNA using primers specific for the product generates a male-specific fragment mapping to the Sxr region. In this way nine potential exons have been positioned on the map. The corresponding cDNAs have not yet been isolated for these fragments but reverse-transcriptase dependent products have been amplified from at least testis RNA for eight of these fragments. These preliminary results indicate that, as suggested by H-Y typing, the Sxrb deletion interval may be relatively complex at the genetic level.

15. Isolation of transcribed sequences from microdissected chromosomal regions

S. Pickering, J. Morten, M. Finnegan, and R. Anand
Zeneca Pharmaceuticals, Cheshire, United Kingdom

We have been evaluating two methods which use chromosomal microdissection for the isolation of transcripts from a region of interest. The first uses DOP amplified microdissected material in direct selection (SHAC, Chen-Lui et al, Genomics 30, 388-392, 1996). The second relies on hybridizing adapter-tagged cDNA to the chromosomes followed by microdissection and recovery of the cDNA by PCR amplification (PrepISH, Hozier et al, Genomics 19, 441-447, 1994, Gracia et al, Hum. Mol. Genet. 5, 595-600, 1996). Both techniques provide a route for the isolation of transcripts from a 5-10 mb region without the need for physical cloning. Most of our work has been on Prep-ISH. To assess the utility of this technique, we have shown that representation is maintained following PCR from high and low copy template. We are currently assessing the degree of enrichment for transcripts from the microdissected region following different hybridization conditions.

16. Isolation and characterization of transcribed sequences from a 1Mbp meningioma candidate region on human chromosome 22q12

Myriam Peyrard
Molecular Medicine, clinical Genetics Unit, Karolinska Hospital, Stockholm, Sweden

Meningiomas are brain tumors arising from the meninges covering the central nervous system. Extensive deletion mapping studies showed that monosomy or loss of large portions of human chromosome 22 are primary events in meningioma tumorigenesis. However, no restricted region of this chromosome arm is commonly deleted; leading to the conclusion that more than one gene on 22q is involved in tumor development.We constructed a cosmid contig covering the largest candidate region (estimated to 1Mbp) in 22q12. At present, as much as 50-60% of this contig is sequenced. In order to clone the transcribed genes contained in this available genomic sequence, we are combining sequence-based exon prediction (using the Grail computer program) and testing of human cDNA libraries. An update in the isolation of transcribed

sequences from this region will be presented.

17. **Transcript mapping in a "gene-rich" region in human chromosome 13q12-q13 containing the breast cancer susceptibility gene BRCA2**

Susan Rhodes(1), Alan Ashworth(2), Laurent Baron(1), Andrew Futreal(3), Pieter de Jong(4), Simon Gregory(1), Gos Micklem(1), Catherine Rice(1), Michael Stratton(2), Richard Wooster(1), and David Bentley(1)
(1)The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, United Kingdom (2)Sections of Molecular Carcinogenesis and CRC Centre for Cell and Molecular Biology, Institute of Cancer Research, Haddow Laboratories, and Chester Beatty Laboratories, United Kingdom (3)Duke University Medical Centre, Durham, North Carolina, USA (4)Human Genetics Department, Rosewell Park Cancer Institute, Buffalo, New York, USA

The availability of genomic sequence provides the means to identify and characterize genes by a combination of computational and experimental techniques. A physical map of a 900 kb region of human chromosome 13q has been constructed using overlapping P1- artificial chromosome clones (PACs). Potential coding sequences were identified from a minimum overlapping set of clones by exon trapping, cDNA selection and direct screening of cDNA libraries. These sequences were added to an analysis from public databases for full-length cDNAs, protein matches and expressed sequence tags (ESTs) across the genomic sequence. All potentially coding sequences were aligned and displayed in ACEDB 4.3.

18. **Functional test of ESTs encoding probable signal sequences for secretion and surface display** Babru Samal
Department of Computational Biology, Amgen, Inc, Thousand Oaks, CA, USA

Few thousand expressed sequence tags have been subjected to a computer program to predict if they encode leader sequences for secreted proteins or receptors. A functional test has been devised to test these predictions. In this assay, the predicted signal sequences substitute for the leader sequence of a reporter gene. COS cells are transfected with the resulting hybrid DNA in a plasmid backbone. Functional substitution of the leader sequence of the reporter by the novel sequence is taken as the confirmation of the computer generated prediction.

19. **Gene mapping at the Canadian Genome Analysis & Technology Program (CGAT) Resource Facility**

S. Scherer(1), J. Huizenga(1), J. Herbrick(1), S. Soder(1), G. Traverso(1), B. Beattie(2), J. Squire(2), C. Wu(3), P. de Jong(3), Y.J. Kang(4), U. Sohn(4) and L.-C. Tsui(1)
(1)Department of Genetics and (2)Molecular Pathology, The Hospital for Sick Children, Toronto, Ontario, M5G 1X8, Canada; (3)Department of Human Genetics, Roswell Park Cancer Institute, Buffalo, NY, USA (4)Department of Genetic Engineering, Kyungpook National University, Taegu, Korea.

Our center was established as a national core facility to provide a service for Canadian investigators wishing to map genes to human chromosomes. Since the mapping reagents provided to us can be of a diverse origin (eg. human or mouse) and type (cDNA, genomic clone or DNA sequence) we have adopted a comprehensive approach in an effort to map the genes with the highest frequency, accuracy, and resolution. Preferably, both gene/cDNA probes and oligonucleotides designed from the DNA sequence are screened by hybridization and PCR, respectively, against the NIGMS somatic cell hybrid mapping panel (#2), genomic libraries (primarily the CEPH mega-YAC library and the RPCI-I total human PAC library), and radiation hybrid panels (Stanford-G3 and GeneBridge-G4 panels). The combination of these techniques usually allows for a precise mapping of the gene within at least one existing framework map. If

it is not possible to determine the cytogenetic position of the gene from these experiments a cDNA clone or corresponding genomic clone is then FISH mapped against metaphase spreads. Using this approach we have successfully mapped over 270 genes across the genome. Since the majority of these genes have been characterized at the functional level the information is immediately useful for positional cloning experiments.

In collaboration with other groups we have also initiated large scale cDNA clone and EST mapping projects. One project involves the mapping of cDNAs and ESTs from a human thymus cDNA sequencing project. Briefly, 1,117 sequences were determined and following vigorous database searches, the DNA sequence of the 3'-UTR of the ESTs not already assigned to a chromosome are being converted into STSs. They are then screened by PCR against a monochromosomal somatic cell hybrid panel and a subset of the CEPH-megaYAC library. To expedite and economize the YAC library screening process we have used the most recent MIT-Whitehead whole genome map and other published maps of single chromosomes (chromosomes 12, 19, 21, and 22) as a guide to select 4000 clones which represent 2-fold coverage tiling paths of each chromosome. The same set of YAC clones is also being used in other projects for rapid screening of retinal and heart specific cDNAs. Progress on our whole-genome mapping project will be reported.

20. **Chemical Cross-linking Subtraction: a subtractive screening procedure that does not require separation or amplification**

Joanne M. Walter(1), Maxine Belfield(1), Ian N. Hampson(2), Christopher A. Read(1). (1)Amersham International plc, Amersham Laboratories, Amersham, United Kingdom (2)Paterson Institute for Cancer Research, Manchester, United Kingdom

Chemical Cross-linking Subtraction (CCLS)(1) is a novel approach to the problem of identifying and isolating differentially expressed genes. Here we describe the method by demonstrating its use in the initial isolation of specifically expressed genes from human foetal brain tissue. 10 micrograms of human heart poly A+ mRNA was hybridized with 400ng foetal brain cDNA in an 80% formamide buffer at 42oC to a Rot of 600. The common species in heart and brain RNA populations hybridized and were inactivated in the subsequent cross-linking step by 2,5 diaziridinyl-1,4-benzoquinone (DZQ). Generation of probes from the single stranded novel species took place using a random primer approach with Sequenase version 2.0 T7 DNA polymerase to prevent labeling of excess RNA and alpha 32P dCTP (Amersham:AA0005). The probe was used to screen approximately 30,000 clones from a directional Lambda MOSSlox foetal brain library and 12 of the potential positives were selected for further analysis by comparison to duplicate screens using unsubtracted heart cDNA. Secondary screening, automatic subcloning and Northern analysis was carried out resulting in 50% of the original primary selected clones showing differential expression. 40% of the clones failed to subclone and 10% were below the detection limits of the Northern analysis used. Having shown the utility of CCLS in this application further work is underway to isolate differentially expressed DNA repair genes after stimulation of a HeLa cell line with UV radiation.

CCLS represents a robust approach to the identification of differentially expressed genes which does not require multiple rounds of subtraction, physical separation of the subtracted cDNA or amplification by PCR. It is likely to find use in a variety of research applications, particularly when a library screening approach is favored over the running of numerous polyacrylamide gels and the system under study enables the generation of adequate levels of RNA.

Reference: HAMPSON, I. N. et al. Nucl. Acids Res., 20, pp 2899,1992.

1996 Roster

Donna Albertson
(1)MRC Lab of Molecular Biology
Hills Road
Cambridge CB2 2QH
United Kingdom
(2)Life Sciences Division
Lawrence Berkeley National
Laboratory
1 Cyclotron Road
Berkeley, CA 94720
USA
415-476-3659
415-476-8218
Donna_Albertson@dmcqm.ucsf.edu

Sandro Banfi
Telethon Inst of Genetics and Medicine
Via Olgettina, 58
20132 Milan
Italy
39-2-215601
39-2-21560220
banfi@tigem.it

Sarah Baxendale
Paterson Institute of Cancer Research
Christie Hospital
Wilmslow Road
Manchester M20 4BX
United Kingdom
44-161-446-3233
44-161-446-3033
exhlp@picr.cr.man.ac.uk

Kevin Becker
Natl Center for Human Genome Research
National Institutes of Health
Bethesda, MD 20893-4470
301-594-5283
301-402-3421
kbecker@nchgr.nih.gov

Guiseppe Borsani
Telethon Inst of Genetics and
Medicine
Via Olgettina, 58
20132 Milan
Italy
39-2-215601
39-2-21560220
borsani@tigem.it
http://office4.tigem.it

Kym Boycott
University of Calgary
Department of Medical Genetics
3330 Hospital Drive N.W.
HMRB257
Calgary, Alberta
T2N 4N1 Canada
403-220-7804
403-283-4740
kboycott@acs.ucalgary.ca

Miles B. Brennan
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80206
USA
303-333-4515
303-333-8423

Anthony J. Brookes
MRC, Human Genetics Unit
Western General Hospital
Crewe Road
Edinburgh, Scotland EH4 2XU
United Kingdom
44-131-343-2620

miles@eri.uchsc.edu

Robert A. Brooksbank
The Sanger Centre
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 1SA
United Kingdom
44-1223-494949
44-1223-494919
rab@sanger.ac.uk
http://www.sanger.ac.uk/~rab/

Pierre Colas
Department of Molecular Biology
Massachusetts General Hospital
Boston, MA 02114
USA
617-726-5956
617-726-6893
Colas@frodo.mgh.harvard.edu

Roger D. Cox
Wellcome Trust Centre for Human
Genetics
Windmill Road
Headington, Oxford OX3 7BN
United Kingdom
44-1865-740-022
44-1865-742-186
roger.cox@well.ox.ac.uk

Sally H. Cross
Edinburgh University
Institute of Cell and Molecular
Biology
Darwin Building, King's Buildings
Mayfield Road
Edinburgh, Scotland EH9 3JR
United Kingdom
44-131-650-8691
44-131-650-5379
Sally.Cross@ed.ac.uk

44-131-343-2620
tony@hgu.mrc.ac.uk

Michael Clinton
Roslin Institute
Edinburgh, Scotland EH25 9PS
United Kingdom
44-131-440-2726
44-131-440-0434
michael.clinton@bbsrc.ac.uk

Martin M. Corcoran
Molecular Biology Laboratory
Department of Pathology
Royal Bournemouth General Hospital
Castle Lane East
Bournemouth BH7 7DW
United Kingdom
44-1202-704805
44-1202-309975
rbhpathology@lds.co.uk

Treasa A. Creavin
University College, London Med School
The Rayne Institute
5 University Street
London WC1E 6JJ
United Kingdom
44-171-2096106
44-171-2096103
rmjjtac@ucl.ac.uk

Laura de Gregorio
Instituto Nazionale Tumori
Via G. Venezian 1
Milan 20133
Italy
39-2-2390642
39-2-2390764
dragani@icil64.cilea.it

Laurent Duret
Medical Biochemistry Department
Geneva University Hospital
1 rue Michel Servet
Geneva CH-1211
Switzerland
41-22-706-9914
41-22-794-6965
duret@dim.hcuge.ch
http://expasy.hcuge.ch/www/lad.html

Greg Elgar
Molecular Genetics
Department of Medicine
Box 157, Level 5
Addenbrookes Hospital
Hills Road
Cambridge CB2 2QQ
United Kingdom
44-1223-402436
44-1223-494512
gelgar@med.cam.ac.uk

John N. Feder
Mercator Genetics
4040 Campbell Avenue
Menlo Park, CA 94025
USA
415-614-7006
415-617-0883
feder@mercator.com

James W. Fickett
SmithKline Beecham Pharmaceuticals
Mailcode UW2230
709 Swedeland Road
King of Prussia, PA 19406
USA
610-270-6242
610-270-5580
ficketjw@molbio.sbphrd.com

Katheleen Gardiner
Eleanor Roosevelt Institute
1899 Gaylord Street
Denver CO 80206
USA
303-333-4515
303-333-8423
gardiner@eri.uchsc.edu

Jaimie Greenham
University College, London Medical School
The Rayne Institute
5 University Street
London WC1E 6JJ
United Kingdom
44-171-209-6107
44-171-209-6103
j.greenham@ucl.ac.uk

Peter C. Groot
University of Utrecht
P.O. Box 80154
3508 TD, Utrecht
The Netherlands
31-30-2531683
31-30-2518126
pcgro@pobox.ruu.nl

Christian Gruber
Life Technologies, Inc.
8717 Grovemont Circle
Gaithersburg, MD 20877
USA
301-670-7714
301-921-2116
cgruber@lifetech.com

Ian N. Hampson
Paterson Institute of Cancer Research
Christie Hospital
Wilmslow Rooad
Manchester M20 4BX
United Kingdom

Lynne Hampson
Paterson Institute of Cancer Research
Christie Hospital
Wilmslow Rooad
Manchester M20 4BX
United Kingdom

44-161-446-3233
44-161-446-3033
exhlp@picr.cr.man.ac.uk

Karl W. Hasel
Digital Gene Technologies, Inc.
11149 N. Torrey Pines Road
La Jolla, CA 92037
USA
619-552-1400
619-552-8625
chuck@dgt.com

Steven J.M. Jones
The Sanger Center
Wellcome Trust Genome Campus
Cambridge CB10 1SA
United Kingdom
44-1223-494976
44-1223-494919
sjj@sanger.ac.uk
http://www.sanger.ac.uk

Bernhard Korn
Deutsches Krebsforschungszentrum
Im Neuenheimer Feld 506
69120 Heidelberg
Germany
49-6221-42-3416
49-6221-42-3454
b.korn@dkfz-heidelberg.de

Wai-Choi Leung
Department of Pathology
Tulane University School of
Medicine
1430 Tulane Avenue
New Orleans, LA 70112
USA
504-588-5237
504-587-7389
wcleung@tmc.tulane.edu

Heather A. McQueen
University of Edinburgh

44-161-446-3233
44-161-446-3033
exhlp@picr.cr.man.ac.uk

Simon Horvat
Roslin Institute
Division of Molecular Biology
Roslin, Scotland EH25 9PS
United Kingdom
44-131-440-2726
44-131-440-0434
simon.horvat@bbsrc.ac.uk

Stephen F. Kingsmore
Center for Mammalian Genetics
University of Florida
P.O. Box 100221
Gainesville, FL 32610-0221
USA
904-392-8600
904-392-8483
kingsmore@cmg.health.ufl.edu.

Yun-Fai Chris Lau
Department of Medicine, VAMC-111C5
University of California, San Francisco
4150 Clement Street
San Francisco, CA 94121
USA
415-476-8839
415-502-1613
clau@itsa.ucsf.edu

Michael McClelland
Sidney Kimmel Cancer Center
30999 Science Park Road
San Diego, CA 92121
USA
619-450-5990
619-550-3998
mcclelland@lifsci.sdsu.edu

Sophie Mazeyrat
INSERM U406

Institute of Cell and Molecular
Biology
Darwin Building, King's Buildings
Mayfield Road
Edinburgh, Scotland EH9 3JR
United Kingdom
44-131-650-5819
44-131-650-5379
Heather.McQueen@ed.ac.uk

Faculte de Medicine
27, Boulevard Jean Moulin
13385 Marseille Cedex 5
France
33-91-25-71-54
33-91-80-43-19
mitchell@ibdm.univ-mrs.fr

Miriam Meisler
Department of Human Genetics
University of Michigan
4708 Medical Sciences II
Ann Arbor, MI 48109-0618
USA
313-763-5546
313-763-9691
meislerm@umich.edu

Michael J. Mitchell
INSERM U406
Faculte de Medicine
27, Boulevard Jean Moulin
13385 Marseille Cedex 5
France
33-91-25-71-54
33-91-80-43-19
mitchell@ibdm.univ-mrs.fr

Douglas Montgomery
Glaxo Wellcome Medicines
Research Centre
Room 6G224
Gunnels Wood Road
Stevenage SG1 2NY
United Kingdom
44-1438-764-950
44-1438-764-231
PS14446@GGR.CO.UK

John E.N. Morten
Zeneca Pharmaceuticals
Mereside, Alderley Park
Cheshire SK10 4TG
United Kingdom
44-1625-514760
44-1625-513441
John.Morten@gbapr.pharmaceuticals.ZENECA.tmailuk.sprint.com

Richard J. Mural
Biology Division
Oak Ridge National Laboratory
P. O. Box 2009
Oak Ridge TN 37831
USA
423-576-2938
423-574-1274
muralrj@ornl.gov

Catherine Nguyen
CIML Luminy
13288 Marseille, Cedex 9
France
33-91-269482
33-91-269430
nguyen@ciml.univ-mrs.fr

Jacob Odeberg
KTH, Royal Institute of Technology
S-100 44 Stockholm
Sweden
46-8-790-87-58
46-8-24-54-52

Patrick Onyango
Research Institute of Molecular Pathology
Dr. BohrGasse 7
A-1030 Vienna
Austria
43-1-79730-423

jacob.odeberg@biochem.kth.se

43-1-798-71-53
Onyango@aimp.una.ac.at

Lucy R. Osborne
The Hospital for Sick Children
Department of Genetics
555 University Avenue
Toronto, Ontario M5G 1X8
Canada
416-813-6365
416-813-4937
lucy@genet.sickkids.on.ca

Marie-Josephe Pebusque
INSERM U119
Laboratoire d'Oncologie Moleculaire
27 bd Lei Roure
13009 Marseille
France
33-91-75-84-07
33-91-26-03-64
pebusque@infobiogen.fr

Myriam Peyrard
Molecular Medicine
Karolinska Hospital
L6-A Building
S-171 76 Stockholm
Sweden
46-8-729-3922
46-8-32-77-34
myp@gen.ks.se

Jennifer Phillips
University of Pennsylvania Medical School
36th and Hamilton Walk
Philadelphia, PA 19104-6084
USA
215-898-0420
215-573-2236
eberwine@mscf.med.upenn.edu

Pierre Pontarotti
INSERM U119
27 bd Lei Roure
13009 marseille
France
33-91-75-84-08
33-91-26-03-64
pontaro@infobiogen.fr

Susan Rhodes
The Sanger Centre
Wellcome Trust Genome Campus
Cambridge CB10 1SA
United Kingdom
44-1223-494842
44-1223-494919
sr1@sanger.ac.uk

Martin Ringwald
The Jackson Laboratory
600 Main Street
Bar Harbor, ME 04609
USA
207-288-6436
207-288-2516
ringwald@informatics.jax.org

Petra Ross-Macdonald
Yale University
KBT 918/Biology
New Haven, CT 06520-8108
USA
203-432-3510
203-432-6161
petra@minerva.cis.yale.edu
http://ycmi.med.yale.edu/YGAC/home.html

Alla Rynditch
1National Academy of Sciences of
Ukraine
Kiev, Ukraine

Babru Samal
Amgen, Inc.
M/S 14-1-D, Amgen Center
Thousand Oaks, CA 91320

2Eleanor Roosevelt Institute
1899 Gaylord Street
Denver, CO 80203
USA
303-333-4515
303-333-8423
gardiner@eri.uchsc.edu

USA
805-447-3063
805-447-1982
bsamal@amgen.com

Stephen W. Scherer
Department of Genetics
The Hospital for Sick Children
Toronto, Ontario M5G 1X8
Canada
416-813-7613
416-813-4931
steve@genet.sickkids.on.ca

Esther Schmidt
Ludwig Institute for Cancer Research
Karolinska Hospital
S-171 76 Stockholm
Sweden
46-8-729-2499
46-8-323662
esther.schmidt@instpat.ks.se

Masahiko Shiraishi
National Cancer Center Research
Institute
1-1, Tskuiji 5-chome
Chuo-ku 104, Tokyo
Japan
81-3-3542-2411
81-3-5565-9535
mshirais@ncc.go.jp

Marcelo Bento Soares
Columbia University
722 West 168th Street
Unit #41
New York, NY 10032
USA
212-960-2313
212-781-3577
cuc@cuccfa.ccc.columbia.edu

Sylvia J. Spengler
Lawrence Berkeley National
Laboratory
1 Cyclotron Road
Berkeley, CA 94720
USA
510-486-4879
510-486-5717

J. Gregor Sutcliffe
The Scripps Research Institute
10666 N. Torrey Pines Rd
La Jolla, CA 92037
USA
619-554-8064
619-554-6612
greg@dgt.com

Jeffrey W. Touchman
National Center for Human Genome
Research
National Institutes of Health
49 Convent Drive, MSC 4431
Bldg 49, Rm 2C36
Bethesda, MD 28092
USA
301-402-4931
301-402-0474
jefft@nchgr.nih.gov

Nitin S. Udar
UCLA School of Medicine
10833 Le Conte Ave
Los Angeles, CA 90095
USA
310-825-7200
310-206-0657
nudar@pathology.medsch.ucla.edu

Esther van de Vosse
Department of Human Genetics
Leiden University
Wassenaarseweg 72
2323 AL Leiden
The Netherlands
31-71-5276105
31-71-5276075
esther@ruly46.medfac.leidenuniv.nl

Joanne M. Walter
Amersham International
White Lion Road
HP7 9LL, Buckinghamshire
United Kingdoom
44-1494-54-3850
44-1494-54-3627
Jo_Walter@amersham.co.uk

Sherman M. Weissman
Yale University School of Medicine
295 Congress Avenue, Rm 336
New Haven CT 06510
USA
203-737-2282
203-737-2286
weissman@biomed.med.yale.edu

Thomas Werner
Forschungszentrum fuer Umweltund Geschundheit
Institute fuer Saeugertiergenetik
Ingolstaedter Landstr. 1
D-85758 Oberschleissheim
Germany
49-89-3187-4050
49-89-3187-4400
werner@gsf.de
http://www.gsf.de/biodv

Gayle E. Woloschak
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439-4833
USA
708-252-3312
708-252-3387
woloschak@anl.gov

Gad Yagil
1Weizmann Institute
Rohovot, Israel
2EBI
Hinxton Hall
Cambridge CB10 1SD
United Kingdom
44-1223-494-403
44-1223-494-471
lcyagil@ebi.ac.uk
lcyagil@wiccmail.weixmann.ac.il

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

ALPHABETICAL LIST OF AUTHORS

Albertson, Donna	Banfi, Sandro	Baxendale, Sarah
Becker, Kevin	Borsani, Guiseppe	Boycott, Kym
Brennan, Miles	Brookes, Anthony	Brooksbank, Robert
Clinton, Michael	Colas, Pierre	Corcoran, Martin
Cox, Roger	Creavin, Treasa	Cross, Sally
de Gregorio, Laura	Duret, Laurent	Elgar, Greg
Feder, John	Fickett, James	Gardiner, Katheleen
Greenham, Jaimie	Groot, Peter	Gruber, Christian
Hampson, Ian	Hasel, Karl	Horvat, Simon
Jones, Steven	Kingsmore, Stephen	Korn, Bernhard
Lau, Yun-Fai Chris	Leung, Wai-Choi	McClelland, Michael
McQueen, Heather	Meisler, Miriam	Mitchell, Michael
Montgomery, Douglas	Morten, John	Mural, Richard
Nguyen, Catherine	Odeberg, Jacob	Onyango, Patrick
Osborne, Lucy	Pebusque, Marie-Josephe	Peyrard, Myriam
Phillips, Jennifer	Pontarotti, Pierre	Rhodes, Susan
Ringwald, Martin	Ross-Macdonald, Petra	Rynditch, Alla
Samal, Babru	Scherer, Stephen	Schmidt, Esther
Shiraishi, Masahiko	Soares, Marcelo Bento	Sutcliffe, J. Gregor
Touchman, Jeffrey	Udar, Nitin	van de Vosse, Esther
Walter, Joanne	Weissman, Sherman	Werner, Thomas
Woloschak, Gayle	Yagil, Gad	Yaspo, Marie-Laure