© 2001 Nature Publishing Group http://genetics.nature.com



gerprint, assembled genomic contig and radiation hybrid maps. We mapped a total of 59,557 UniGene clusters on the basis of at least two independent criteria, compared with the 30,000 human genes mapped in Genemap '99. The extension of the human transcript consensus in this study enabled a grater number of putative functional assignments than the 12,700 annotated entries in UniGene. This study reports a first attempt at mapping and annotating a majority of the human transcript into the human genome draft. This information can be immediately applied to the discovery of new genes and the identification of candidate genes that either directly cause or increase our susceptibility to neoplastic diseases.

*Zarbl, Helmut* [71]

## Patterns of gene expression in mammary cells following carcinogen exposure of strains of rat sensitive or resistant to carcinogenesis

Helmut Zarbl, Andrei Mikheev & Robert Sullivan

*Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA*

We previously demonstrated that the activating *Hras1* oncogene mutations found in rat mammary tumors induced by *N*-nitroso-*N*-methylurea (NMU) arise as background mutations within cells of the developing gland. We further showed that NMU enhanced the phenotypic penetrance of these pre-existing mutations by initiating alterations in DNA conformation within the promoter region of *Hras1* and presumably other target genes. The observed epigenetic response to NMU was present in strains of rats that are genetically predisposed to mammary cancers (such as F344), whereas strains of rats resistant to mammary carcinogenesis (such as Cop) did not show this epigenetic response. Genetic linkage studies by Gould and co-workers have demonstrated that the differential sensitivities to mammary carcinogenesis among rat strains are due to the presence of one or more putative, cell-autonomous, mammary carcinoma suppressor genes in the resistant strains. Other studies have demonstrated that differential sensitivities among rat strains are not due to differences in the rate of DNA adduct formation or removal after exposure. We are using rat GeneChip (Affymetrix) technology in an effort to understand the biochemical basis for this altered response to carcinogen. By comparing the patterns of gene expression as a function of time after exposure to NMU in the resistant and sensitive strains (F344 × Cop)F1 and N2 backcross progeny, we hope to identify the biochemical pathways affected by loss of the putative Cop tumor suppressor gene. Preliminary findings on differences in response to mammary carcinogen exposures will be presented.

*Zhao, Lue Ping* [72]

## Toward rigorous assessment of statistical significance in functional genomics

Lue Ping Zhao & Jeff Thomas

*Fred Hutchinson Cancer Research Center, Seattle, Washington, USA*

In the analysis of thousands of genes with only a small number of samples, a major challenge facing functional genomics is the assessment of statistical significance, that is, determining how many findings are false or the estimated rate of false-positive results. We describe a simple permutation method to evaluate exactly the genome-wide significance level, acknowledging both multiple comparisons and small sample sizes. We illustrate this method through an application to a study of leukemia.

*Chen, Jianjun* [73]

## Genome-wide analysis of gene expression in myeloid progenitor cells

Jianjun Chen, Sanggyu Lee, Janet D. Rowley & San Ming Wang

*University of Chicago, 5841 Maryland Avenue, Chicago, Illinois 60637, USA*

Cells in the hematopoietic system undergo extensive phenotypic changes during the differentiation process, starting from stem cells, entering the committed lineage, and then to maturation. Each of these events must be tightly controlled by expression of specific sets of genes in the genome. Various genetic alternations, for example, those in leukemia will change the normal pattern of gene expression and will lead to the pathological transformation. Therefore, more complete data on the pattern of gene expression in both normal and leukemic states will likely result in a better understanding of the genetic control of hematopoietic differentiation and will provide clues for understanding the cause of abnormal differentiation.

We have initiated studies on genome-wide gene expression in both normal progenitor and leukemic cells. We have developed an integrated system that includes four "wet-bench" techniques and a bioinformatic program for this project: 1)The SPGI technique (Screening PolydA/dT(-) cDNAs for Gene Identification); 2) The IPGI technique (Integrated Procedures for Gene Identification); 3) A modified SAGE technique (Serial Analysis of Gene Expression); 4) the GLGI technique (Generation of Longer cDNA fragments from SAGE tags for Gene Identification).

Our goals are: 1). Qualitatively identify all the genes expressed in the targeted cells; 2). Quantitatively analyze the difference in the gene expression pattern between the normal and leukemic cells; 3). Identify the critical pathways through which altered gene expression leads to leukemia with the goal of using this information to develop targeted therapeutic strategies. 4). Applying the whole system to investigate the genome, for example, to identify the correct number of genes.

Over 27,000 and 22,000 unique tag sequences technique from human and mouse myeloid progenitor cells ($CD15^+$ and $GR\text{-}1^+$) have been identified respectively by applying SAGE method. Analysis of these unique tag sequences in the both species showed: 1) over 70% genes are expressed at low level whereas a small number of tags were expressed at high and intermediate levels; 2) about 60% of tags matched to known expressed sequences and 40% of tags had no match and these represent potentially novel genes; 3) analysis of the highly expressed genes identified the genes specifically expressed in myeloid progenitor cells, as well as genes ubiquitously expressed in other cell types. The data from this study provide a base for analyzing the genome-wide gene expression in the differentiation of human and mouse myeloid progenitor cells. The data also provide a baseline for further characterization of abnormal gene expression in the myeloid lineage in pathological states, especially in leukemia.



© 2001 Nature Publishing Group http://genetics.nature.com