significant Pfam hits, 4,520 sequences were in the cancer subclass. At present 59 Interpro entries exist in this subclass. These entries include breast cancer susceptibility proteins, retinoblastoma protein domains, p53 tumor antigen, Xeroderma pigmentosum proteins and the Burkitt's lymphoma receptor.

1. Apweiler R. *et al. Nucl. Acids Res.* **29**, 37–40 (2001).

Yang, Ping

[67]

Neutrophil elastase gene in lung cancer development: evidence from molecular genetics and clinical epidemiology

Ping Yang, Ken Taniguchi, Claude Deschamps, Eric Bass, Rebecca Meyer & Wanguo Liu

Mayo Foundation, Rochester, Minnesota 55901, USA

Neutrophil elastase (NE) is a powerful protease capable of degrading various protein components of the extracellular matrix, coagulation and component cascades. Local production of NE is involved in invasion and associated with a poor lung cancer prognosis. We propose that NE is important in lung cancer development, especially in the invasive phase. We tested this hypothesis using 344 cases and 299 controls. We detected single-nucleotide polymorphisms at the NE locus by denaturing high-performance liquid chromatography analysis, DNA sequencing and cloning of the NE gene promoter region. We performed transfection analysis, using relative luciferase activity on a human lung cancer cell line, to determine the regulatory role of the detected single-nucleotide polymorphisms in setting transcription levels of NE promoter. Two new single-nucleotide polymorphism markers, REP_A (T and G) and REP_B2 (G and A), were detected. There was no GG at REP_A and very little AA at REP_B2. TT and TG were contrasting allele types at REP_A, and GG and AA/AG were contrasting allele types at REP_B2. Measured by odds ratio, the TT type at REP_A or GG type at REP_B2 was associated with a 2.3 or 1.4 times higher risk of developing lung cancer than the TG and AA/AG types, respectively. When assessing the combined effects of the high-risk alleles, the odds ratio was 24.8. We demonstrated a 1.9-fold increase of relative luciferase activity in the T-G construct compared with the G-A construct, providing evidence that the TT-GG type correlates with a high NE level. Our findings indicate an important role for NE in lung cancer development.

Yang, Xiang-Jiao

[68]

Monocytic leukemia zinc-finger protein MOZ and its related factor MORF are new histone acetyltransferases

Nathalie Champagne, Nadine Pelletier & Xiang-Jiao Yang

Molecular Oncology Group, McGill University Health Centre, Montreal, Quebec H3A 1G4, Canada

The monocytic leukemia zinc-finger protein (MOZ) gene is rearranged in t(8;16)(p11;p13) and inv(8)(p11q13) associated with acute myeloid leukemia. The other fusion partners involved are CBP and TIF2, both of which are known transcriptional coactivators with intrinsic histone acetyltransferase activity. We have cloned MORF, a new human protein related to MOZ, and demonstrated that MOZ and MORF have intrinsic histone acetyltransferase activity. Moreover, like MORF, MOZ possesses a weak transcriptional repression domain at its N-terminal part and a strong activation domain at its C-terminal part. These results indicate that MOZ and MORF are acetyltransferases involved in regulating transcription and thereby shed light on how aberrant MOZ proteins lead to leukemogenesis.

Yeung, Ka Yee

[69]

Transcriptional analysis of Barrett's epithelium and normal gastrointestinal tissues

Ka Yee Yeung¹, Michael Barrett², Jeff Delrow², Patricia Blount², Brian Reid² & Peter Rabinovitch³

¹Department of Computer Science and Engineering, University of Washington, Seattle, Washington, USA

²Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

³Department of Pathology, University of Washington, Seattle, Washington, USA

Barrett's esophagus is a premalignant condition caused by chronic acid reflux in which the normal squamous epithelium of the esophagus is replaced by a metaplastic columnar epithelium. Of interest is the distinction between neoplastic Barrett's epithelium (BE) and surrounding normal tissues of the upper gastrointestinal tract. For example, although it arises in the esophagus, BE more closely resembles the epithelium of the duodenum at the histological level. We compared the transcriptional profile of BE to the profiles of normal upper gastrointestinal tissues, including gastric epithelium, squamous epithelium of the esophagus and duodenal epithelium. We collected endoscopic biopsies from each tissue from a series of patients during routine surveillance. Poly(A)⁺ RNA was prepared from pooled samples (2–4 patients per pool) of BE (four pools), esophageal squamous epithelium (four pools), gastric epithelium (three pools) and duodenal epithelium (three pools) and used to interrogate Affymetrix HU6800 and FL6800 chips. We found no difference in the correlation coefficients between neoplastic BE and each of the normal tissues. In addition, we searched for tissue-specific patterns of gene expression in the normal tissues and the neoplastic BE. We compared the performance of several clustering algorithms. Our results suggest that the data set consists of approximately eight clusters, and the best performance was obtained using CAST and k-means. From this analysis, we applied CAST to identify eight clusters, among which some were specific for squamous epithelium (203 genes), duodenal epithelium (211 genes), gastric epithelium (105 genes) and BE (36 genes).

Yuan, Bo

[70]

Physical mapping and functional annotation of 60,000 human genes

Degen Zhuo¹, Wei Zhao¹, Hee-Yung Yang², Jian-Ping Wang¹, Russell Sears¹, Do-Hun Kwon¹, David Gordon¹, Solomon Gibbs¹, Dai Dean², Troy Baer³, Don Stredney³, Al Stutz³, Ralf Krahe¹, Fred Wright¹ & Bo Yuan¹

¹Bioinformatics Group, Human Cancer Genetics Program, Ohio State University, Columbus, Ohio, USA

²Labbook.com, Columbus, Ohio, USA

³Ohio Supercomputer Center, Columbus, Ohio, USA

The recent release of the first draft of the human genome provides an unprecedented opportunity to integrate all human genes and their functions within a complete positional context. However, at least four significant technical hurdles remain: to create a complete and nonredundant human transcript index, to assemble the still-fragmented human genome draft, to place the individual transcript indices accurately on the human genome and to annotate all human genes functionally. We report the extension of the UniGene database through the assembly of its sequence clusters into nonredundant sequence contigs. The resulting consensus was aligned to the draft genome. We determined a unique location for each transcript within the human genome by the integration of the restriction fin-

Abstracts: Session III

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 greater 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-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.