한국가금학회지 제 28권 제 2호, 143~153 (2001)
Korean J. Poult. Sci. Vol. 28, No. 2, 143~153 (2001)



# Discovering Novel Genes of Poultry in Genomic Era

S. K. Kang, B. C. Lee, J. M. Lim[1], J. Y. Han[1] and W. S. Hwang

*Department of Theriogenology & Biotechnology, College of Veterinary Medicine, Seoul National University, Seoul 152-742, Korea*

*[1] School of Agricultural Biotechnology, Seoul National University, Suweon 441-744, Korea*

***ABSTRACT :*** Using bioinformatic tools for searching the massive genome databases, it is possible to identify new genes in few minutes for initial discoveries based on evolutionary conservation, domain homology, and tissue expression patterns, followed by further verification and characterization using the bench-top works. The development of high-density two-dimensional arrays has allowed the analysis of the expression of thousands of genes simultaneously in the humans, mice, rats, yeast, and bacteria to elucidate the genes and pathways involved in physiological processes. In addition, rapid and automated protein identification is being achieved by searching protein and nucleotide sequence databases directly with data generated from mass spectrometry. Recently, analysis at the biochemical level such as biochemical screening and metabolic profiling (Biochemical genomics) has been introduced as an additional approach for categorical assignment of gene function. To take advantage of recent achievements in computational approaches for facilitated gene discoveries in the avian model, chicken expression sequence tags (ESTs) have been reported and deposited in the international databases. By searching EST databases, a chicken heparanase gene was identified and functionally confirmed by subsequent experiments. Using combination of subtractive hybridization assay and Genbank database searches, a chicken heme-binding protein family (cSOUL/HBP) was isolated in the retina and pineal gland of domestic chicken and verified by Northern blot analysis. Microarrays have identified several host genes whose expression levels are elevated following infection of chicken embryo fibroblasts (CEF) with Marek's disease virus (MDV). The ongoing process of chicken genome projects and new discoveries and breakthroughs in genomics and proteomics will no doubt reveal new and exciting information and advances in the avian research.

(Key words: gene discovery, genomics, proteomics, poultry)

## INTRODUCTION

A project with the goal of determining the complete nucleotide sequence of the human genome was first formally proposed in 1985 (1). The Human Genome Project (HGP) in 1990 was officially initiated in the United States under the direction of the National Institutes of Health and the U.S. Department of Energy. In 1998, Celera also announced plan for completing the human genome sequence. At the present time, nearly complete genomic sequences of 2.91-billion base pair (bp) consensus sequence of the euchromatic portion of the human genome was revealed (2, 3). In addition, complete genomic sequences of more than two dozens organisms including yeast *Saccharomyces cerevisiae* and nematode *Caenorhabditis elegans* are already available. This progress became possible through the largely automated sequencing of libraries of genomic DNA and expressed sequence tags (ESTs) that represent fragments of transcribed genes from diverse tissues. The ESTs provide a rapid and reliable method for gene discovery as well as a resource for the large-scale analysis of gene expression pattern of known and unknown genes at specific stages of development and in specific tissues (4, 5). These genome sequences are now being used as a frame-

work to allow the identification of novel genes and mutations involved in diseases, and investigation of biological processes such as metabolic pathways, signal networks, functional classes and protein fold types by the aid of intensified computational approaches. As genome sequences are nearly completed, the field of functional genomics seeks to devise and apply technologies that utilize the massive body of sequence information to achieve the full identification of novel genes and encoded proteins simultaneously, while traditional methodologies study a small number of gene at a time. Such new genomic and proteomic technologies are aimed at (a) discovering new genes and proteins, (b) quantifying and analyzing gene and protein expression, and (c) assigning functionality.

## NOVEL GENE DISCOVERIES BY COMPUTATIONAL GLOBAL APPROACHES

Paradigm shift in research of gene discovery currently is underway. Traditional studies of identification of novel genes are based upon protein purification and protein-protein interaction assays (6). Following the advent of recombinant DNA technology, the new genes have been discovered by using low-stringency hybridization, degenerative PCR, differential cloning (subtraction and differential display) (7, 8), and positional cloning following chromosome walking (9). The searches for interacting protein were achieved by expression cloning methods based on ligand or antibody binding (10, 11, 12) and yeast two-hybrid interaction assay (13, 14, 15). These approaches have been effective but labor-intensive for initial gene discovery and linking function of proteins to genes. With near completion of genome sequences and major advances in bioinformatic search tools available on the Internet (Table 1), these genomic and technological revolution have prompted the emergence of new research paradigms for the identification of novel genes and proteins through easy accesses to massive resources to genome and protein databases. Utilizing the unprecedented power of modern computing tools, the new approaches take few minutes for initial discoveries of new genes compared to time-consuming the laboratory works. Furthermore, performing changes the expectation or threshold values in computational searches can replace the labor-intensive and tedious stringency control procedures in the traditional cloning methods based on sequence matching (e.g. degenerate PCR and low-stringency hybridization).

## DATABASE MINING FOR NOVEL GENES BY HOMOLOGY SEARCHES

The use of databases to determine sequence similarities of gene products has facilitated the discovery of novel homologs in various organisms. This idea was based on the concept that genes are present in families with conservation in structure and function in diverse organisms through the evolution of simple organisms into complex ones (16). To achieve efficient comparison of DNA or protein sequences, a variety of paired sequence comparison programs using different scoring matrices have been developed for aligning individual sequences with a catalogued sequence database. For example, the Basic Local Alignment Search Tool (BLAST) and other related programs are being used for gene sequence analysis and the deduction of their functions (17, 18, 19) (Table 1). Using the computational searches in the GenBank, several G protein-coupled receptors have been identified based on sequence conservation in the transmembrane region (20, 21). Four novel leucine-rich repeat-containing, G protein-coupled receptors (LGR) were isolated from *Drosophila* and snail *Lymnanei stagnalis* using primitive homologous sequences from vertebrate glycoprotein receptors as queries for the GenBank search (22, 23). A number of genes homologous to mammalian glycoprotein hormone receptors were identified from nematode by applying the same evolutionary conservation approach (24). Similar bioinformatic searches have identified multiple cytokines (25, 26, 27, 28) and a large group of human Toll-like recep-

**Table 1.** Bioinformatic tools and useful web links for new gene discovery

| Name | Description/Features | Web Site |
|---|---|---|
| The EST Machine | Directory of links to selected bioinformatic resources for EST analysis | http://www.tigem.it/ESTmachine.html |
| GENSCAN | Predict genes in genomic DNA based on probabilistic models in gene structure | http://CCR-081.mit.edu/GENESCAN.html |
| NetGene | Predict splice sites in vertebrate genes | http://130.225.67.199/services/NetGene/index. |
| NetGene | Predict splice sites in vertebrate genes | http://130.225.67.199/services/NetGene/index.html |
| Genie | Identification of multi-exon genes in *Drosophila* genomic sequences | http://www.fruitfly.org/seq_tools/genie.html |
| FGENES | Predict genes and exons by using pattern-based structure | http://genomic.sanger.ac.uk/gf/gf.html |
| ExPASy Molecular Biology Server | Links to databases, tools, and other molecular biology resources | Http://www.expasy.ch/ |
| PROSITE | A database of protein families and domains | Http://www.expasy.ch/prosite/ |
| Pfam | Classification of predicted proteins into protein domain families | http://pfam.wustl.edu/ |
| MEME | Detect motifs (conserved sequence patterns) by statistical technique | http://www.sdsc.edu/MEME/meme/website/intro.html |
| Blocks WWW Server | A service for biological sequence analysis by pattern-searching method or statistical method | http://www.blocks.fhcrc.org/blocks/ |
| PRINTS (Protein Fingerprint Database) | A compendium of protein fingerprints | http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/PRINTS.html |
| eMOTIFs | Database of profile and sequence motifs for representing structural and functional aspects of proteins. Forms motifs or subsets of aligned sequence | http://dna.stanford.edu/identify/ |
| SMART | A simple modular architecture research tool for the identification of signaling domains | http://smart.embl-heidelberg.de/ |
| UniGene | A nonredundant set of gene-oriented clusters. | http://www.ncbi.nlm.nih.gov/Web/Newsltr/aug96.html#advance |
| STACK | Comprehensive representation of the sequence of each of the expressed genes in the human genome | http://www.sanbi.ac.za/Dbases.html |
| TIGR | A collection of databases containing DNA and protein sequences, gene expressions, cellular roles, protein families, and taxonomic data | http://www.tigr.org/tdb/index.html |
| Transfac | A relational database of transcriptional factor binding site and transcription factors | http://transfac.gbf.de/TRANSFAC |

**Table 1.** Continued

| Name | Description/Features | Web Site |
|---|---|---|
| TESS | Finds potential transcription-factor-binding sites in DNA sequence | http://cbil.upenn.edu/tess |
| TSSG | Recognizes human Pol II promoter region and transcription initiation sites | http://genomic.sanger.ac.uk/gf/HELP/tssw.html |
| GRAIL | Identification of multi-exon genes | http://avalon.epm.ornl.gov/Grail-in/EmptyGrailForm |
| Polyadq | Identification of a true poly(A) signal | http://sciclio.cshl.org/mzhanglab/abaska/polyadq_form.html |
| Genotator | A tool package runs a series of sequence analysis tools and displayed color-coded sequence annotations in a browser | http://www.fruitfly.org/~nomi/genotator-paper.html |
| NJ tree | Molecular phylogenetic tree by neighbor-joining method | http://www.biophys.kyoto-u.ac.jp/maketree2.html |
| SignalP | Prediction of signal peptide sequence | http://www.cbs.dtu.dk/services/SignalP/ |
| PSORT | Prediction of protein sorting signals and localization sites in amino acid sequences | http://psort.nibb.ac.jp/ |
| COILS | Prediction of coiled coil regions in proteins | http://www.ch.embnet.org/software/COILS_form.html |
| TMHMM | Prediction of transmembrane helices in proteins | http://www.cbs.dtu.dk/services/TMHMM-1.0/ |
| SCOP | Classification of protein database for analysis of sequences and structures | http://scop.mrc-lmb.cam.ac.uk/scop/ |
| TopPred2 | Topology prediction of membrane proteins | http://www.sbc.su.se/~erikw/toppred2/ |
| DNA motif and structure tools | Collection of links to databases, tools, and other molecular biology resources | http://dapsas.weizmann.ac.il/bio_tools/dna-tools.html#orf |

tors (29, 30, 31). Alternatively, if limited significant homology is observed for query sequence using the similarity methods, the sequence can be scanned by other methods which search for shorter regions of conservation that represent conserved sequence motifs (32). These searches employ tools such as eMOTIFs, BLOCKs and PRINTS. Taking advantage of this computational approach, two novel relaxin/insulin-like factors (RIFs) sharing significant homology with relaxin in the putative mature portion of these polypeptides were identified (33). Based on domain conservation, GenBank searches have led to isolation of an intracellular apoptosis mediator DEFT (Death EFfector in Testis) that is abundantly expressed in the testis (34). The structure of a gene and protein identified by homology searches can be predicted using tools such as FGENES, Pfam, SMART and TMHMM (Table 1).

## COMPREHENSIVE GENE-EXPRESSION PROFILING

Gene-expression profiling, which monitors the expression of thousands of genes simultaneously, can be accomplished by several techniques including

cDNA arrays, oligonucleotide arrays, serial analysis of gene expression (SAGE) and differential display. The development of high-density two-dimensional arrays of chemically synthesized molecules on the various foundations (cDNA arrays, oligonucletide arrays, electrokinetic arrays and fiberoptic arrays) has allowed the analysis of the expression of thousands of genes simultaneously (35). Chip-based quantitative RNA expression experiments are being applied to the study of humans, mice, rats, yeast, and bacteria to understand the genes and pathways involved in physiological processes. For example, DNA chips designed for analyzing the expression levels of >6,000 genes in *S. cerevisiae* identified more than 400 RNAs whose levels are significantly changed as the cell progresses through the cell cycle (36). High-density arrays have been used to identify yeast genes whose expression depends on transcriptional initiation factors, to profile gene expression changes following activation of mouse T cells (37), and to explore and compare signal transduction pathway (38, 39). Also, DNA arrays have been employed to identify human genes involved in the pathology of diseases such as rheumatoid arthritis and inflammatory bowel disease (40), to compare gene expression in cells expressing either a transformed or a nontransformed phenotype (41, 42), and to study hematopoietic differentiation (43). In combination with cluster analysis, microarrays identified variation in gene expression patterns of human cancers as a means to classify solid tumors (44). Microarray analysis is also widely recognized as a key tool in drug discovery (45) and is being used to characterize human genetic variation (single nucleotide polymorphisms, or SNPs) (46). Extensive stretches of DNA sequence can be screened at once, and more than 4,000 common genetic variations have been found across the human genome (46). These small differences can be useful markers in subsequent studies to identify the genes responsible for particular traits and to analyze the sequences of important genes to uncover predisposition to common disease or the efficacy and safety of therapies. In general, the bench-top works such as Northern hybridization for selected genes are used to confirm gene expression changes first observed on microarrays.

## COMPREHENSIVE PROTEIN DISCOVERY USING MASS SPECTROMETRY AND COMPUTATIONAL SEARCHES

The elucidation of an organism's genome and the data created by whole-genome sequencing has stimulated the development of proteomics by providing a sequence infrastructure for the identification of novel proteins and their functions. The term proteomics describes the ability to apply global (proteom-wide or system-wide) experimental approaches to isolate and assess protein function (47). Proteomics has emerged as a new experimental approach because in part mass spectrometry has simplified protein analysis and characterization. In addition, several important recent innovations in protein separations using high performance liquid chromatography (HPLC) and capillary electrophoresis (CE), along with recent advances in mass spectrometry instrumentation, display technology, biochips, imaging, and automation have revolutionized proteomics (48, 49, 50, 51, 52). Rapid and automated protein identification can be achieved by searching protein and nucleotide sequence databases directly with data generated from two-dimensional gel electrophoresis analysis, followed by peptide mass fingerprinting with MALDI-TOF (matrix-assisted laser-desorption-ionization-time-of-flight) mass spectrometry. An alternative approach based on the proteolytic digestion of protein mixtures, which is followed by reversed-phased liquid chromatography to separate or partially fractionate the complex peptide mixture and direct introduction into a tandem mass spectrometer, has been developed (47). This approach reduces the reliance on SDS-PAGE to separate proteins, provides a more flexible strategy for proteolytic digestion and manipulation, and can enhance the sensitivity of mass spectrometry. Using such approach, at least ten proteins which were not observed by 2-D gel electrophoresis were isolated in

the yeast ribosome (53). Furthermore, modern mass spectrometers can provide sufficient information to allow unique recognition of protein fragments as well as detection of secondary modifications such as phosphorylation and glycosylation. Also, integration of a genome database with protein coprecipitation and global yeast two-hybrid interaction approaches to enrich proteins of interest should allow one to assemble comprehensive interaction maps of genomes (54, 55) and for routine identification of downstream components of cell surface receptors in mammalian cells (56).

## GENE AND FUNCTIONAL PROTEIN DISCOVERIES VIA METABOLIC PROFILING

Although undoubtedly very powerful, expression arrays and mass spectrometry approaches has limitations that they do not necessarily pinpoint the function of a gene (57). For example, a change in mRNA levels observed in the expression array does not necessarily mean that the corresponding gene actually has a function within the biological process. Information from expression arrays always needs to be confirmed by further experimentation before an assignment of function of genes. Similarly, at the proteomic level, even though higher levels of a protein are detected, it does not necessarily mean that more activity will be found within the structural, metabolic or signaling network associated with this protein. Therefore, analysis at the biochemical level such as biochemical screening and metabolic profiling (Biochemical genomics) was introduced as an additional approach for categorical assignment of gene function. Using *in vitro* biochemical screening approach, Martzen et al. demonstrated that tRNA ligase and 2'-phosphodiesterase, two known yeast tRNA splicing activities, were detected only in the pools known to contain the respective GST-ORF fusion proteins (58). Metabolic profiling technologies are also being used to investigate the metabolic consequence of a particular mutation or targeted alteration of gene activity in plant and mammalian systems. Sophisticated gas chromatography-mass spectrometry techniques detected around 150 compounds in developing potato tubers (59), probed the metabolism of Arabidopsis leaves and distinguished some 326 distinct compounds (60). Analysis for volatile organic compounds by gas chromatography-mass spectrometry in liquid and gaseous samples from human urine has led to identification of 34 compounds (59). Profiling of bile acids, cholesterol, and triglycerides formed as a part of a metabolic study conducted on mice that lacked a bile-acid receptor (61). In this study, combinational approaches in histology, gene-expression profiling and biochemical profiling contributed to a better understanding of how altered gene function can influence the cell structure and metabolic function of an organism (61).

## DISCOVERING NOVEL GENES OF POULTRY IN GENOMIC ERA

The chicken genetic map consists of about 235 genes-based markers or about 12% of the total number of DNA markers (62). To take advantage of recent achievements in computational approaches for facilitated gene discoveries in the avian model, several chicken ESTs such as an activated T-cell ESTs (63), bursal lymphocytes ESTs (64) and a white Leghorn chicken embryo ESTs (65) have been reported and deposited in the international databases. Currently, there are two publicly available chicken EST databases (http://www.chickest.udel.edu; http://genetics.hpi.uni-hamburg.de). By searching EST databases with the cDNA sequence of human heparanase as a query sequence, a chicken heparanase gene was identified and functionally confirmed by subsequent experiments (66). Using combination of subtractive hybridization and Genbank database searches (nr; via blastx and blatn), a chicken heme-binding protein family (cSOUL/HBP) was isolated in the retina and pineal gland of domestic chicken, *Gallus gallus*, and verified by Northern blot analysis (67). Morgan et al. utilized Microarrays containing 1,126 nonredundant cDNAs

selected from a chicken activated T-cell expressed sequence tag database (63) to investigate changes of host cell gene expression that accompany infection of chicken embryo fibroblasts (CEF) with Marek's disease virus (MDV) (68). Elevated host genes by MDV vial infections include macrophage inflammatory protein, interferon response factor 1, interferon-inducible protein, quiescence-specific protein, thymic shared antigen 1, major histocompatibility complex (MHC) class I, MHC class II, ?2-microglobulin, clusterin, interleukin-13 receptor alpha chain, ovotransferrin, a serine/threonine kinase, and avian leucosis virus subgroup J glycoprotein.

## CONCLUSION

We live in an age of gene discovery. As genome projects of chicken progress and the availability of near complete genome sequences of many species including human, the massive archives in the databases represent a golden opportunity to discover novel genes and proteins. The current shift in the research from one gene at a time to a global approach has led emergence of new methods to integrate the explosion of knowledge on gene sequences, transcript expression profiles, and protein functions and interactions.

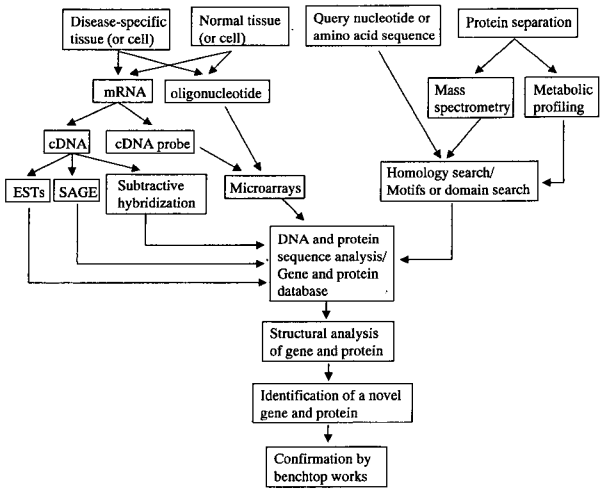


**Fig. 1.** The working flow of funcional genomic and proteomic.

Through comparative sequence analysis, expression arrays, mass spectrometry and metabolic profiling, novel chicken genes and their functions can be revealed in next few years. The working flow of functional genomics and proteomics was described in Fig. 1. The ongoing discoveries and breakthroughs are genomic and proteomics will no doubt reveal new and exciting information and advances in the avian research.

## ACKNOWLEDGMENTS

This work was supported by grants from the Korean Ministry of Science and Technology (G7 project; #98-G-08-02-A-03) and the Brain Korea 21 project (#0503-20000007)

## REFERENCES

Sinsheimer RL 1989 Sequencing the human genome: Summary Report of the Santa Fe Workshop. Genomics 5:1989.

Aach J, Bulyk ML, Church GM, Comander J, Derti A, Shendure J 2001 Computational comparison of two draft sequences of the human genome. Nature 409:856-859.

Venter JC et al 2001 The sequence of the human genome. Science 291:1304-1351.

Boguski MS, Schuler GD 1995 Establishing a human transcript map. Nat Genet 10:369-371.

Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tome P, Aggarwal A, Bajorek E, Bentolila S, Birren BB, Butler A, Castle AB, Chiannilkulchai N, Chu A, Clee C, Cowles S, Day PJ, Dibling T, Drouot N, Dunham I, Duprat S, East C, Hudson TJ, et al. 1996 A gene map of the human genome. Science 274:540-546.

Margolis B 1994 The GRB family of SH2 domain proteins. Prog Biophys Mol Biol 62:223-244.

Tsuchida K, Mathews LS, Vale WW 1993 Cloning and characterization of a transmembrane serine kinase

that acts as an activin type I receptor. Proc Natl Acad Sci USA 90:11242-11246.

Koenig BB, Cook JS, Wolsing DH, Ting J, Tiesman JP, Correa PE, Olson CA, Pecquet AL, Ventura F, Grant RA, et al. 1994 Characterization and cloning of a receptor for BMP-2 and BMP-4 from NIH 3T3 cells. Mol Cell Biol 14:5961-5974.

Chandrasekharappa SC, Guru SC, Manickam P, Olufemi SE, Collins FS, Emmert-Buck MR, Debelenko LV, Zhuang Z, Lubensky IA, Liotta LA, Crabtree JS, Wang Y, Roe BA, Weisemann J, Boguski MS, Agarwal SK, Kester MB, Kim YS, Heppner C, Dong Q, Spiegel AM, Burns AL, Marx SJ 1997 Positional cloning of the gene for multiple endocrine neoplasia-type 1. Science 276:404-407.

Mathews LS, Vale WW 1991 Expression cloning of an activin receptor, a predicted transmembrane serine kinase. Cell 65:973-982.

Takemoto Y, Furuta M, Sato M, Hashimoto Y 1997 A simple improvement in expression cloning. DNA Cell Biol 16:797-799.

Franzen P, ten Dijke P, Ichijo H, Yamashita H, Schulz P, Heldin CH, Miyazono K 1993 Cloning of a TGF beta type I receptor that forms a heteromeric complex with the TGF beta type II receptor. Cell 75:681-692.

McLaughlin MM, Kumar S, McDonnell PC, Van Horn S, Lee JC, Livi GP, Young PR 1996 Identification of mitogen-activated protein (MAP) kinase-activated protein kinase-3, a novel substrate of CSBP p38 MAP kinase. J Biol Chem 271:8488-8492.

Witczak O, Skalhegg BS, Keryer G, Bornens M, Tasken K, Jahnsen T, Orstavik S 1999 Cloning and characterization of a cDNA encoding an A-kinase anchoring protein located in the centrosome, AKAP450. Embo J 18:1858-1868.

Kawabata M, Chytil A, Moses HL 1995 Cloning of a novel type II serine/threonine kinase receptor through interaction with the type I transforming growth factor-beta receptor. J Biol Chem 270: 5625-5630.

Bruccoleri RE, Dougherty TJ, Davison DB 1998 Concordance analysis of microbial genomes. Nucleic Acids Res 26:4482-4486.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ 1990 Basic local alignment search tool. J Mol Biol 215:403-410.

Altschul SF, Lipman DJ 1990 Protein database searches for multiple alignments. Proc Natl Acad Sci USA 87:5509-5513.

Pearson WR, Lipman DJ 1988 Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444-2448.

Marchese A, George SR, Kolakowski LF, Jr., Lynch KR, O'Dowd BF 1999 Novel GPCRs and their endogenous ligands: expanding the boundaries of physiology and pharmacology. Trends Pharmacol Sci 20:370-375.

O'Dowd BF, Nguyen T, Jung BP, Marchese A, Cheng R, Heng HH, Kolakowski LF, Jr., Lynch KR, George SR 1997 Cloning and chromosomal mapping of four putative novel human G-protein- coupled receptor genes. Gene 187:75-81.

Tensen CP, Van Kesteren ER, Planta RJ, Cox KJ, Burke JF, van Heerikhuizen H, Vreugdenhil E 1994 A G protein-coupled receptor with low density lipoprotein-binding motifs suggests a role for lipoproteins in G-linked signal transduction. Proc Natl Acad Sci USA 91:4816-4820.

Hauser F, Nothacker HP, Grimmelikhuijzen CJ 1997 Molecular cloning, genomic organization, and developmental regulation of a novel receptor from Drosophila melanogaster structurally related to members of the thyroid-stimulating hormone, follicle-stimulating hormone, luteinizing hormone/ choriogonadotropin receptor family from mammals. J Biol Chem 272:1002-1010.

Kudo M, Chen T, Nakabayashi K, Hsu SY, Hsueh AJ 2000 The nematode leucine-rich repeat-containing, G protein-coupled receptor (LGR) protein homologous to vertebrate gonadotropin and thyrotropin receptors is constitutively actived in mammalian cells. Mol Endocrinol 14:272-284.

Wells TN, Peitsch MC 1997 The chemokine information source: identification and characterization of novel chemokines using the WorldWideWeb and

expressed sequence tag databases. J Leukoc Biol 61:545-550.

Yoshie O, Imai T, Nomiyama H 1997 Novel lymphocyte-specific CC chemokines and their receptors. J Leukoc Biol 62:634-644.

Hieshima K, Imai T, Opdenakker G, Van Damme J, Kusuda J, Tei H, Sakaki Y, Takatsuki K, Miura R, Yoshie O, Nomiyama H 1997 Molecular cloning of a novel human CC chemokine liver and activation-regulated chemokine (LARC) expressed in liver. Chemotactic activity for lymphocytes and gene localization on chromosome 2. J Biol Chem 272: 5846-5853.

Yang JY, Spanaus KS, Widmer U 2000 Cloning, Characterization and Genomic Organization of Lcc-1 (Scya16), a Novel Human Cc Chemokine Expressed in Liver. Cytokine 12:101-109.

Rock FL, Hardiman G, Timans JC, Kastelein RA, Bazan JF 1998 A family of human receptors structurally related to Drosophila Toll. Proc Natl Acad Sci USA 95:588-593.

Chaudhary PM, Ferguson C, Nguyen V, Nguyen O, Massa HF, Eby M, Jasmin A, Trask BJ, Hood L, Nelson PS 1998 Cloning and characterization of two Toll/Interleukin-1 receptor-like genes TIL3 and TIL4: evidence for a multi-gene receptor family in humans. Blood 91:4020-4027.

Yang RB, Mark MR, Gray A, Huang A, Xie MH, Zhang M, Goddard A, Wood WI, Gurney AL, Godowski PJ 1998 Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. Nature 395: 284-288.

Brutlag DL 1998 Genomics and computational molecular biology Curr Opin Microbiol 1: 340-345.

Hsu SY 1999 Cloning of two novel mammalian paralogs of relaxin/insulin family proteins and their expression in testis and kidney. Mol Endocrinol 13: 2163-2174.

Leo CP, Hsu SY, McGee EA, Salanova M, Hsueh AJ 1998 DEFT, a novel death effector domain-containing molecule predominantly expressed in testicular germ cells. Endocrinology 139:4839-4848.

Cunningham MJ 2000 Genomics and proteomics: The new millennium of drug discovery and development. J Pharmacol Toxicol Methods 44:291-300.

Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman D, Lockhart DJ, Davis RW 1998 A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 2:65-73.

Teague TK, Hildeman D, Kedl RM, Mitchell T, Rees W, Schaefer BC, Bender J, Kappler J, Marrack P 1999 Activation changes the spectrum but not the diversity of genes expressed by T cells. Proc Natl Acad Sci USA 96:12691-12696.

Fambrough D, McClure D, Kazlauskas A, Lander ES 1999 Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. Cell 97:727-741.

Madhani HD, Galitski T, Lander ES, Fink GR 1999 Effectors of a developmental mitogen-activated protein kinase cascade revealed by expression signatures of signaling mutants. Proc Natl Acad Sci USA 96:12530-12535.

Heller RA, Schena M, Chai A, Shalon D, Bedoin T, Gilmore J, Woolley DE, Davis RW 1997 Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc Natl Acad Sci USA 94:2150-2155.

DeRisi JL, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM 1996 Use of a cDNA microarray to analyze gene expression patterns in human cancer. Nat Genet 14:457-460.

Trent JM, Bittner M, Zhang J, Wiltshire R, Ray M, Su Y, Garcia E, Meltzer P, DeRisi J, Penland L, Brown P 1997 Use of microgenomic technology for analysis of alterations in DNA copy number and gene expression in malignant melanoma. Clin Exp Immunol 107(Suppl 1):33-40.

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR 1999 Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA 96:2907-2912.

Perou CM, Jeffrey SS, Van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JCF, Lashkari D, Shalon D, Brown PO, Botstein D 1999 Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci USA 96:9212-9217.

Gray NS, Wodicka L, Thunnissen AM, Norman TC, Kwon S, Espinoza FH, Morgan DO, Barnes G, LeClerc S, Meijer L, Kim SH, Lockhart DJ, Schultz PG 1998 Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. Science 281:533-538.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES 1998 Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 281:533-538.

Yates JR 3rd 2000 Mass spectrometry from genomics to proteomics. TIG 16:5-8.

Bradbury J 2000 Proteomics: the next step after genomics? Lancet 356:50.

Li M 2000 Applications of display technology in protein analysis. Nat Biotechnol 18:1251-1256.

MacBeath G, Schreiber SL 2000 Printing proteins as microarrays for high-throughput function determination. Science 289:1760-1762.

Pandey A, Mann M 2000 Proteomics to study genes and genomes. Nature 405:837-845.

Patterson SD 2000 Proteomics: the industrialization of protein chemistry. Curr Opin Biotechnol 11:413-418.

Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd 1999 Direct analysis of protein complexes using mass spectrometry. Nat Biotech 17: 676-682.

Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M 2000 Protein interaction mapping in *C. elegans* using proteins involved in vulval development. Science 287:116-122.

Blackstock WP, Weir MP 1999 Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol 17:121-127.

Pandey A, Podtelejnikov AV, Blagoev B, Bustelo XR, Mann M, Lodish HF 2000 Analysis of receptor signaling pathways by mass spectrometry: Identification of Vav-2 as a substrate of the epidermal and platelet- derived growth factor receptors. Proc Natl Acad Sci USA 97:179-184.

Trethewey RN 2001 Gene discovery via metabolic profiling. Curr Opinon Biotechn 12:135-138.

Martzen MR, McCraith SM, Spinelli SL, Torres FM, Fields S, Grayhack EJ, Phizicky EM 1999 A biochemical genomics approach for identifying genes by the activity of their products. Science 286 :1153-1155.

Roessner U, Wagner C, Kopka J, Trethewey R.N, Willmitzer L 2000 Simultaneous analysis of metabolites in potato tuber by gas chromatography - mass spectrometry. Plant J 23:131-142.

Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L 2000 Metabolic profiling represents a novel and powerful approach for plant functional genomics. Nat Biotechnol 18:1157-1161.

Trethewey RN, Riesmeier JW, Willmitzer L, Stitt M, Geigenberger P 1999 Tuber specific expression of a yeast invertase and a bacterial glucokinase in potato leads to an activation of sucrose phosphate synthase and the creation of a sucrose futile cycle. Planta 208:227-238.

Groenen MAM, Crooijmans RPM, Veenendaal A, Cheng HH, Siwek M, and Van der Poel JJ 1998 A comprehensive microsatellite linkage map of the chicken genome. Genomics 49:265-274.

Tirunagaru VG, Sofer L, Cui J, Burnside J 2000 An expressed sequence tag database of T-cell-enriched activated chicken splenocytes: sequence analysis of 5251 clones. Genomics 66:144-151.

Abdrakhmanov I, Lodygin D, Geroth P, Arakawa H, Law A, Plachy J, Korn B, Buerstedde JM 2000 A large database of chicken bursal ESTs as a

resource for the analysis of vertebrate gene function. Genome Res 10:2062-2069.

Smith EJ, Shi L, Drummond P, Rodriguez L, Hamilton R, Ramlal S, Smith G, Pierce K, Foster J 2001 Expressed sequence tags for the chicken genome from a normalized 10-day-old White Leghorn whole embryo cDNA library: 1. DNA sequence characterization and linkage analysis. J Hered 92 :1-8.

Goldshmidt O, Zcharia E, Aingorn H, Guatta-Rangini Z, Atzmon R, Michal I, Pecker I, Mitrani E, Vlodavsky I 2001 Secretion and expression pattern of human and chicken heparanases are determined by their signal peptide sequence. J Biol Chem, in press.

Zylka MJ, Reppert SM 1999 Discovery of a putative heme-binding protein family (SOUL/HBP) by two-tissue suppression subtractive hybridization and database searches. Mol Brain Res 74:175-181.

Morgan RW, Sofer L, Anderson AS, Bernberg EL, Cui J, Burnside J 2001 Induction of host gene expression following infection of chicken embryo fibroblasts with oncogenic Marek' s disease virus. J Virol 75: 533-539.