

“A System Biology” Approach to Bioinformatics and Functional Genomics in Complex Human Diseases: Arthritis

M.G. Attur¹, M.N. Dave¹, K. Tsunoyama⁴, M. Akamatsu⁴, M. Kobori⁴, J. Miki⁴, S.B. Abramson^{1,2}, M. Katoh⁴, and A.R. Amin^{*1,2,3}

¹Departments of Rheumatology and Medicine, Hospital for Joint Diseases, New York, NY 10003, USA

²Departments of Pathology and Medicine, New York University School of Medicine, New York, NY 10016, USA

³Kaplan Cancer Center, New York, NY 10016

⁴Molecular Medicine Laboratories, Yamanouchi Pharmaceutical Co. Ltd., Tsukuba, Ibaraki, Japan

*Corresponding author: Rheumatology Research and Laboratory for Functional Genomics, Hospital for Joint Diseases/ NYU School of Medicine, 301 East 17th Street, Rm. 1600, New York, NY 10003, USA

Abstract

Human and other annotated genome sequences have facilitated generation of vast amounts of correlative data, from human/animal genetics, normal and disease-affected tissues from complex diseases such as arthritis using gene/protein chips and SNP analysis. These data sets include genes/proteins whose functions are partially known at the cellular level or may be completely unknown (e.g. ESTs). Thus, genomic research has transformed molecular biology from “data poor” to “data rich” science, allowing further division into subpopulations of subcellular fractions, which are often given an “-omic” suffix. These disciplines have to converge at a systemic level to examine the structure and dynamics of cellular and organismal function.

The challenge of characterizing ESTs linked to complex diseases is like interpreting sharp images on a blurred background and therefore requires a multi-dimensional screen for functional genomics (“functionomics”) in tissues, mice and zebra fish model, which intertwines various approaches and readouts to study development and homeostasis of a system. In summary, the post-genomic era of functionomics will facilitate to narrow the bridge between correlative data and causative data by quaint hypothesis-driven research using a system approach integrating “intercoms” of interacting and interdependent disciplines forming a unified whole as described in this review for Arthritis.

Introduction

Arthritis is a complex disease with an unknown etiology. Based on the clinical symptoms, it can be classified as Osteoarthritis, Rheumatoid Arthritis, Synovial Lipomatosis, Avascular Necrosis, Crystal Deposition Disease, Gout, and other diseases. The common underlining symptoms in the above clinical manifestations include inflammation, destruction (of cartilage and soft tissue) and dysfunction of joints (McCarty, 1998).

At the dawn of the new millennium, the major challenge we face is characterization of genes involved in oligo- and polygenic disorders, such as arthritis because unlike monogenic diseases, pedigrees from complex diseases reveal no Mendelian inheritance patterns and gene mutations are neither sufficient nor necessary to explain the disease phenotypes. Several genes regulate many of the diagnostic features of these complex diseases, called Quantitative Trait Locus (QTL) disorders, (Doerge, 2002). For example, Perola *et al* reported QTL analysis of stature (*i.e.* height) from genome scans of five Finnish study groups (Perola *et al.*, 2001). QTLs affecting stature were observed on chromosome 7pter and 9q. Regulation of QTL by several genes is not a rule of thumb because in invertebrate species positional cloning of a QTL revealed that a single gene was responsible for the complex regulation of tomato fruit size (Frery *et al.*, 2000). It is possible that some of the human QTLs may also prove to carry a single gene responsible for complex disorders. Recently preliminary maps of QTLs in heterogeneous stocks of mice have been investigated, which may facilitate structure-functional analysis of similar genes in man (Masinde *et al.*, 2001; Mott and Flint, 2002).

The knowledge of new genomic information and the tools to decipher it, puts us back to square one in our continuing saga to determine the etiology and pathogenesis of joint destruction in arthritis. This obviates the necessity to reassess our working hypothesis. For the first time, the “genomic tools” will allow us to analyze small amounts of surgical samples (such as needle biopsies) and clinical samples in the context of the whole genome, which we have never done before. Preliminary genomic analysis in osteoarthritis (OA) has already resurrected the debate on osteoarthritis or osteoarthrosis based on the semantic issues in the definition of inflammation in the post-genomic era of molecular medicine (Attur *et al.*, 2002a). Further analyses will not only facilitate development of unbiased hypotheses at the molecular level, but also assist us in following the scent to the identification and characterization of novel targets and disease markers for pharmacological intervention, gene therapy and diagnosis.

*For correspondence. Email amina01@popmail.med.nyu.edu; Tel. 212-598-6537; Fax. 212-598-7604.

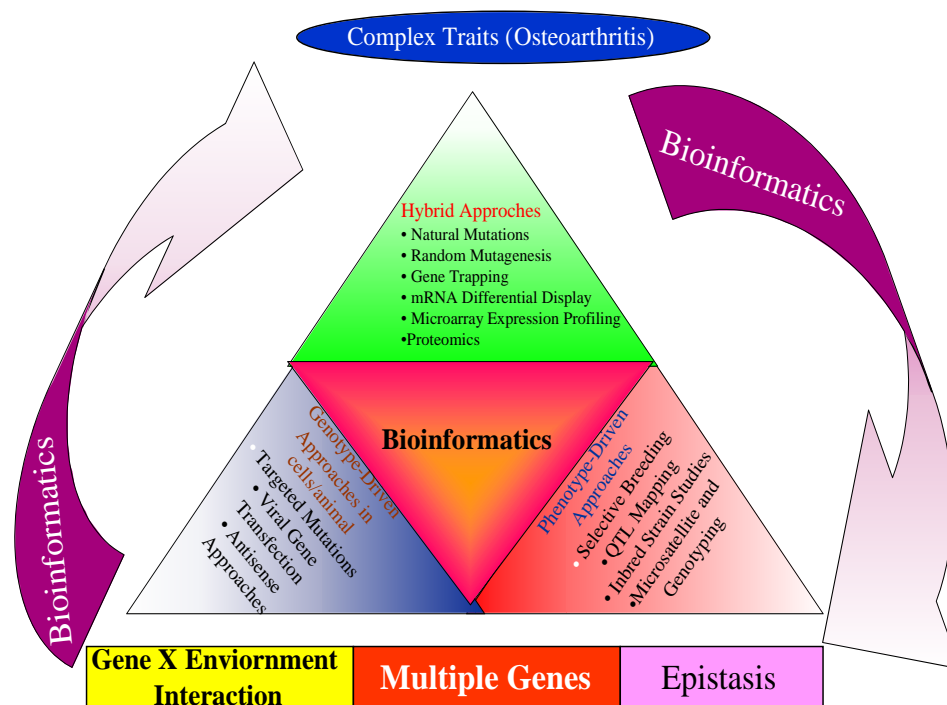


Figure 1. A System Biology Approach to Genomics to study Complex Human Diseases. Multiple genes influence complex traits in complex human disease and require different technologies and expertise to identify and characterize them. At least four independent approaches need to be taken to unravel this complex interaction. (a) Phenotype-driven approaches, where one starts with the trait and traces it to the gene(s) of influence, as described for CACP and PPD in this review. (b) Genotype-driven approaches, wherein the starting point is a specific gene that is traced to a phenotype. This strategy is becoming less distinct as the full sequence of the human and several other model organism's genomes are near completion with annotation. Among these, mice have contributed immensely to pinpointing genetic locations (mapping) and identifying disease-associated genes (Rozzo *et al.*, 2001). Zebra fish models show enormous promise as simple "*in vivo* models" for functionomics to validate the genotypes. (c) Hybrid approaches, where the starting point could be either the gene or the phenotype. This approach revolves around development of technologies such as gene expression arrays. For example, historically it was the case that their phenotypic effects identified natural mutations and traced from the phenotype to the gene (phenotype-driven). However, once all the SNPs and QTLs are identified, mapped and annotated, it will develop a road map to identifying naturally occurring mutations in specific genes and potentially tracing their phenotypic effects by combining it with gene expression arrays and bioinformatics. (d) Bioinformatics: This is a link and "tape" that crystallizes these approaches into a "whole-istic Biology". It has become almost a prerequisite for all genomic approaches.

It is also clear that identifying genetic interactions with environmental conditions and characterizing gene-gene interactions (epistasis) will play an important role in ultimately describing the genetic architecture of complex traits in man. Parallel studies in inbred disease specific mouse strains in controlled environment and genetic makeup will facilitate paving the pathways to understanding these complex interactions.

The present review discusses a system approach to gene mining, bioinformatics, data validation, and functional genomics using arthritis as a complex disease model. We are now simultaneously confronting the complexity of the human genome sequence, complex diseases regulated by multiple genes and risk factors and multiple technological approaches provided by multiple interdisciplinary experts. Furthermore, other challenges in performing genomic research on human subjects and biological material include informed consents, public acceptance, sample collection and storage (Greely, 2001; Magill, 2002).

A System Approach to Arthritis

Arthritis is a disease with complex traits influenced by various risk factors (Brandt *et al.*, 2001; Silman, 2002). Such diseases with multiple genetic, environmental and epistatic

determinants represent the greatest challenge for genetic analysis largely due to the difficulty of isolating the phenotype of one gene amid the noise of other genetic and environmental influences. Unraveling the genetics of human diseases such as arthritis will require moving beyond the focus on one gene at a time to exploring pleiotropism, epistasis and environmental-dependency of genetic effects by integrating various technologies and data sets forming a unified whole. There is consensus among various investigators that single genetic approach is not sufficient to give a comprehensive analysis of a complex disease but rather, would require an entire arsenal of approaches simultaneously as shown in Figure 1 will be required simultaneously. We believe that the combination of classical approaches with the modern genomics approaches will rapidly advance the field.

Gene Mining in Arthritis

Basically, there are several approaches to RNA and protein-based gene mining in complex diseases (Amin, 2000; Attur *et al.*, 2002b,e; Strohmaan, 2002). Isolation of RNA from different biological samples for gene expression analysis reflects the expectations or objectives of the study. The fundamental types of samples commonly used for gene expression analysis include those derived from *in vivo* sources such as postmortem dissections or biopsies of a specific target tissue and those derived from tissue sections using laser capture microscopy. *In vivo* samples can be extremely complex because RNA and/or proteins are derived from many different cell types (intentionally or inadvertently) and the microarray or proteomics data reflects the contribution of all cell types present. It may be impossible to delineate the contribution of different cell

types to a given mRNA or protein sample, yet this sample type is often preferred as it represents the complex processes underlying the biology of a particular disease or other clinically relevant phenotype.

In some cases, the clinical sample from a homogenous cell population may be available in limiting amounts. Pooling samples is an alternative strategy for analyzing such rare and limited samples, although individual data of samples such as clinical observations will be lost in the process. Pooling clinical samples has worked well in our hands primarily because the disease-associated genes are stratified in the pooled samples (Figure 2). At least Real Time PCR using other pooled samples validated 55% of these. The 45% error may be due to the variation of gene expression in both patients and normal people even in pooled samples, and the limitation of the chip technology. These concepts have also been used in arthritis and other

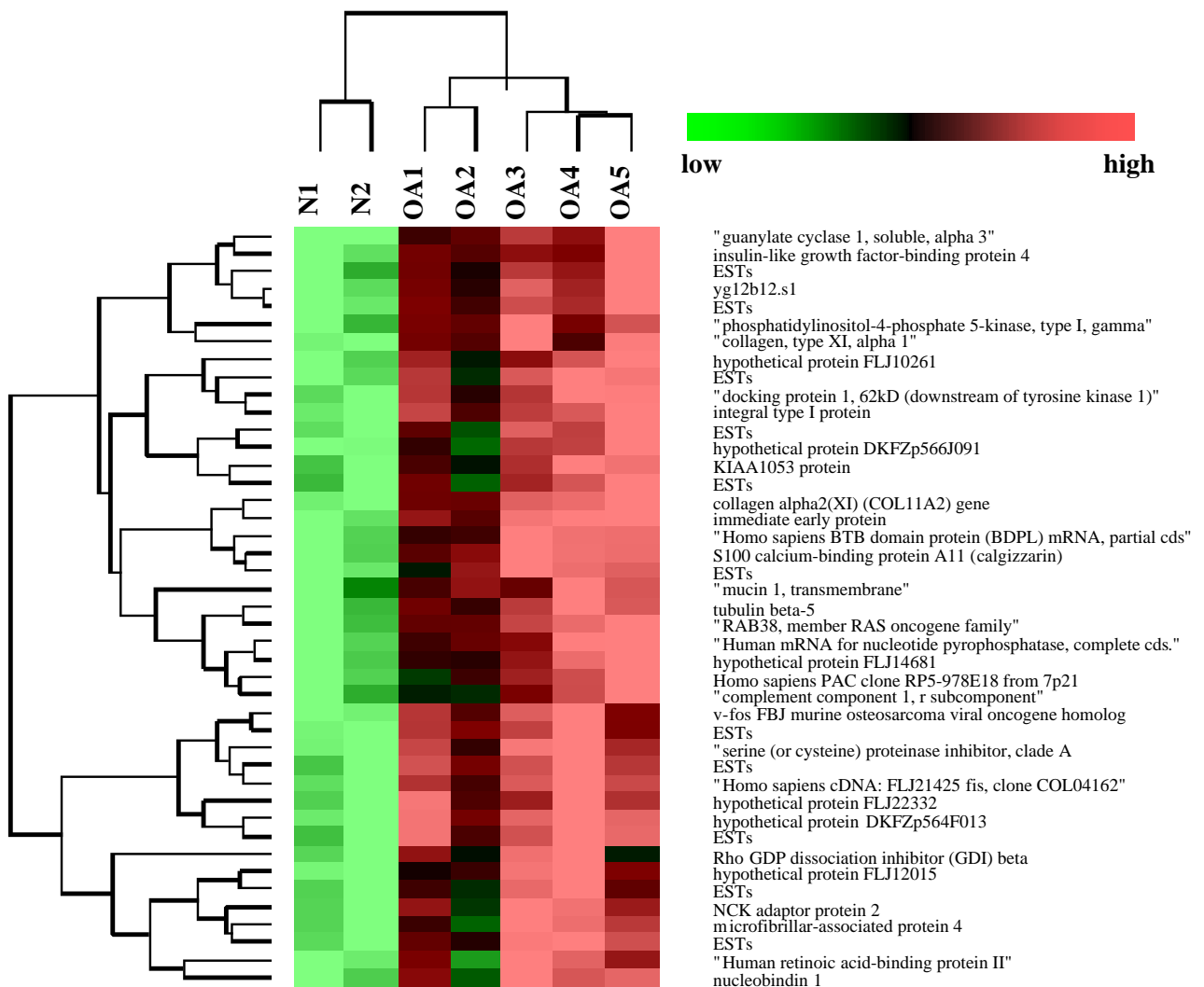


Figure 2. Gene Expression profiling in normal and OA-affected cartilage. Expression array was performed using an Affymetrix gene chip. Two pools of normal (n=20) and five pools of OA cartilage (n=70) samples were utilized. Heat map and hierarchical clustering analysis of 43 genes and ESTs were selected for representation. Gene expression profiles are shown in rows. Red indicates that the gene is expressed more (2 to 10 fold) as compared to basal levels shown in green. The selected genes represent group 1, 2 or 3 based on the criteria described in Figure 3.

diseases (Aigner *et al.*, 2001; Agrawal *et al.*, 2002).

It is estimated that 10,000 to 20,000 transcripts and 50,000 to 80,000 proteins are expressed in a given cell at a certain time period (Lander *et al.*, 2001; Venter *et al.*, 2001). Therefore, gene mining efforts, either individually or in combination, have the tendency to generate an enormous number of differentially expressed transcripts, EST tags and proteins, which may be involved (either directly or indirectly) in the disease process. These potential targets have to be reduced to a manageable number by compiling the ones that may be involved in the disease process by prioritizing several criteria and chosen references. Computational data analysis and clustering algorithms as described below can be optimized for each project to extract relevant information. Another strategy to reduce targets to a manageable number would be by applying pharmacogenomic screens (*in vitro* and *in vivo*), and comparing the targets to chemogenomic databases during pharmacological intervention *e.g.* TNF α and/or recombinant TNF antagonists (Attur *et al.*, 2000a).

Bioinformatics

Bioinformatics is a science, which aims to derive new biological knowledge from various kinds of biological data in terms of molecules, genes, cells, and organisms by applying information technologies. The combined use of mathematics, statistics, and information science enables biologists to understand and organize the biological information on a large scale (Luscombe *et al.*, 2001).

In the post-genomic era, biological data is being produced at a phenomenal rate. An experimental laboratory can produce over 100 gigabytes of data a day with ease. As a result of this surge in data production, computers have become an indispensable tool even for biologists analyzing results of single experiments. In particular, genomics research needs massive computer power to organize, compile, and decipher the complex dynamics observed in biological systems. Bioinformatics has emerged as a key area to address the computational needs of genomics research.

Basically, the aims of bioinformatics are three-fold. (I) The first aim is to compile biological data. Bioinformatics helps researchers to produce biological data through Laboratory Information Management System (LIMS), and to store data in the database. In order to access the stored information, submit new queries and formulate new hypotheses, it is a prerequisite to compile a database. Compilation includes rectifying human sequence data and array information in a systematic manner. It also requires updating the existing information, which comes from various types of the other databases such as DNA/protein sequence database, tertiary structure database, pathway or interaction database, literature database, and so forth. By compilation of the database, bioinformatics boosts the value of the database and provides speculations, which lead to new biological knowledge. (II) The second aim is to develop tools for analyses of the data. In order to achieve this aim, mathematics, statistics, and information science technologies have been applied to the biological data. In addition, distinctive tools of biology have been also

developed. As a simple example, sequence similarity search programs such as BLAST and FASTA (Bottomley, 1999) use the substitution matrices of amino acids for evaluation of similarity among amino acid sequences. These matrices are based on the observation of actual amino acid sequence data and have been developed along with accumulation of sequence data (Bottomley, 1999). Development of bioinformatics tools requires expertise in computational theory, as well as thorough understanding of biology. Moreover, not only tools for analysis of the data but also tools to visualize the results of analyses are indispensable especially in genomics research. Gaining the perspective of huge amount of data is also useful for understanding biology. (III) The third aim is to analyze the data by using various tools and computational power, and to interpret the results in a biologically meaningful manner. In bioinformatics, we can now conduct global analyses of all the available data with the aim of uncovering common principles that apply across many systems and highlight novel features.

Gene Expression Arrays: Normalization, Data Analysis, and Bioinformatics

Microarray experiments in particular have raised a wide range of computational requirements, including image processing, instrumentation and robotics, database design based on available expressed sequence tags (ESTs), and data analysis, data storage and retrieval. Furthermore, microarray data need to be interpreted in the context of other biological knowledge, involving various types of post-genomics informatics, including gene networks, gene pathways, and gene ontologies (Wu, 2001; Quackenbush, 2001).

Affymetrix microarray chips are routinely used in our laboratory for global gene expression studies. The hybridization signal has been shown to be proportional to actual transcript levels based on parallel studies performed using Real Time PCR with identical RNA samples (Attur *et al.*, 2002a). Additionally, the technology has been described as capable of distinguishing concentration levels within a factor of 2, and of detecting transcript frequencies as low as 1 in 2,000,000. The above technology is capable of detecting as little as 100 pM of RNA. Given that a significant number of genes of biological interest have transcript frequencies as low as 1 pM (Chudin *et al.*, 2001; Mahadevappa and Warrington, 1999), the commercial usefulness of this technology is constrained by the minimum abundance level that is reliably detectable. A linear correlation between signal and transcript abundance was consistently observed for transcript concentrations between 1-10 pM. The signal is not linear between 10 pM to 80 pM and becomes saturated after 80 pM. Indeed, the Affymetrix chip array was able to detect low transcript level (0.5 to 1.5 pM) in the absence of a significant background. The results are not reliable with 0.1 pM of transcripts. It is possible to argue that post-hybridization amplification would improve detection, but obviously at the expense of potentially saturating expression levels of more abundant genes. Perhaps scanning images before and after amplification could

maximize detection without suffering saturation penalties. Longer hybridization cycles seem to be a viable alternative, as these enabled partial detection of transcripts (about 5 out of 15) at the 0.1 pM level (Chudin *et al.*, 2002). Affymetrix stated that setting the lower intensity level with PMT (photomultiplier tube) voltage change of scanner could reduce the saturation without missing minimum detectable range. Pushing the envelope may facilitate identifying varying transcripts but could simultaneously distort the general RNA expression profiles. In summary, the data generated for low abundant transcripts is extremely variable and requires a second level of validation.

Normalization of Data

There are four widely used approaches to normalize gene expression data generated using microarrays. All of these are based on the assumption that an exogenous control has been spiked into the RNA before labeling it. The normalization factor that is obtained from spiking (positive control) is then adjusted with the data to compensate for the experimental variability.

Total intensity normalization data relies on the assumption that the quantity of initial mRNA is the same for both labeled samples that are compared. Under this assumption, a normalization factor can be calculated and used to re-scale the intensity for each gene in the array (Quackenbush, 2001).

Normalization using regression analysis is the second approach. For mRNA derived from closely related samples, a significant fraction of the assayed genes would be expected to be expressed at similar levels. For example, in a scatter plot of Cy5 versus Cy3 intensities, these genes would cluster along a straight line. The slope of which would be one if the labeling and detection efficiencies were the same for both samples. In many experiments, the intensities are nonlinear, and local regression techniques are more suitable, such as LOWESS regression (Quackenbush, 2001).

The third approach is normalization using ratio statistics (Chen *et al.*, 1997). The authors assume that although individual genes might be up or downregulated, in closely related cells, the total quantity of RNA produced is approximately the same for essential genes such as housekeeping genes. Using this assumption, they developed an approximate probability density for the rationale. They then describe how this can be used in an iterative process that normalizes the mean expression ratio to one and calculates confidence limits that can be used to identify differentially expressed genes.

The fourth normalization strategy used the intensities of house keeping genes where expression of 7,000 full-length genes in eleven different human tissues was examined (Warrington *et al.*, 2000). The authors predicted that 535 transcripts, which could serve as likely candidates for housekeeping genes. Forty-seven of these were consistently and commonly expressed between adult and fetal samples and could serve as housekeeping genes in developmental biology. Housekeeping genes in normal and diseased tissues have to be analyzed on a case to case basis in normal and diseased tissue before a judgment

can be made. For example house keeping transcripts, GAPDH, acidic ribosomal protein, β -actin, cyclophilin, phosphoglycerokinae, β 2-microglobulin, β -glucosidase, hypoxanthine ribosyl transfrase and transferin receptor were analyzed by microarray in normal and OA-affected cartilage. Among these, GAPDH and acidic ribosomal protein were expressed at relatively higher level and showed consistent expression in normal and diseases cartilage. (Amin, unpublished data).

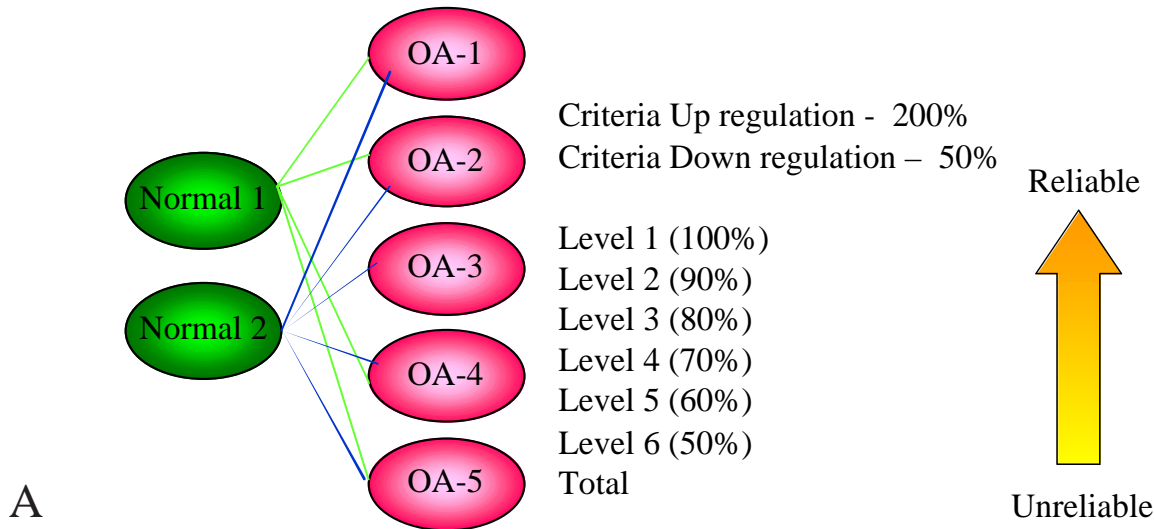
Cluster Analysis

Gene expression analysis generates significant amounts of data. To interpret the results from such multiple data sets, it is helpful to have an intuitive visual representation. Programs have been designed to switch data generally by reordering the rows/columns or both, such that, patterns of expression become visually apparent when presented in this junction. In this regard, the cluster analysis, which is one of the classical statistic methods, is most frequently used. Applying this method to gene expression data can group together genes with similar expression patterns and also can categorize samples with similar expression profiles.

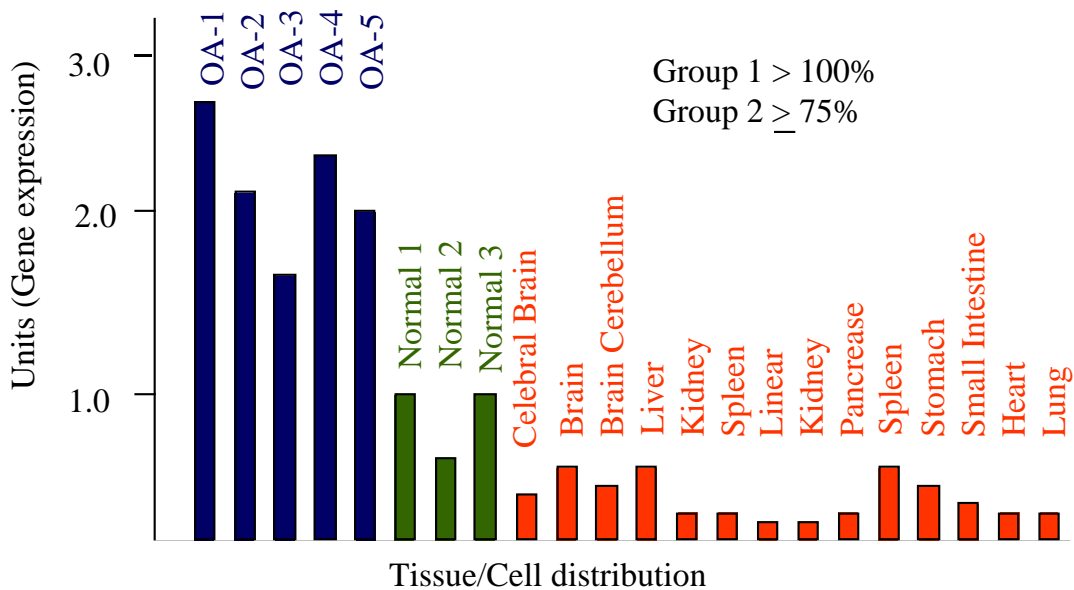
In general, clustering methods are divided into hierarchical and non-hierarchical methods. As for hierarchical clustering methods, there are several algorithms, which differ in the manner of distances among genes or clusters and the manner of constructing clusters. In the calculation of distances, if necessary, adequate transformation of expression values is required, such as the logarithmic transformation or the normalization in which expression values of each gene or sample have mean = 0 and variance = 1 (as distances, correlation coefficients and Euclidean distances are widely used). The algorithms of constructing clusters include, but are not limited to a) single linkage method, b) complete linkage method, c) unweighted pair-group average method, d) centroid method, and e) the Ward's method. The result of these hierarchical clustering methods is described as a dendrogram. Hierarchical clustering methods have been noted by statisticians to have the problem of lacking robustness and complicating interpretation of the hierarchy. In order to avoid these problems, non-hierarchical methods can be used. For instance, self-organizing maps is one of the non-hierarchical methods which is suitable and effective for microarray data analysis (Quackenbush, 2001). The choice of the methods or algorithms described above may be determined by "robustness of clusters" or reasonableness to biological interpretation. Thus, in order to obtain right conclusion, it is prudent to examine several methods and weigh the results.

Genome-Wide Scans

We all share at least 99.9% of the nucleotide code in our genome. Yet, it is remarkable that the diversity encoded by less than 0.1% variation in our DNA represents almost all the diverse phenotypes seen in man. These diverse phenotypes also include genes susceptible to complex diseases. These can be analyzed by genome-wide scans,



A



B

Figure 3. Identification of osteoarthritis and cartilage-specific genes in human OA.

(A) Classification of Osteoarthritis associated genes based on their representation.

The up- and down-regulated genes in OA-affected cartilage were defined as transcripts that were upregulated by 200%, or decreased by less than 50% in OA cartilage as compared to normal cartilage, respectively. The gene expression profiles of 2 normal pools (n=20) and 5 OA pools (n=70) were compared in 10 different combinations as shown in the figure. The reliability of OA associated genes can be judged on the number of comparisons satisfied by these criteria. The most reliable genes satisfied with these criteria in 10 out of 10 comparisons, and were classified as level 1 of OA associated genes. The genes satisfying the criteria in nine, eight, seven, and six of 10 comparisons were classified as level 2, 3, 4, and 5, respectively. Other genes revealed up- and down-regulation in less than five comparisons were excluded because of their lower reliability. In summary, 1,469 genes in total were characterized as OA associated genes.

(B) Tissue distribution of OA associated genes.

The Gene Chip data of OA associated genes were compared with that of 14 normal tissues using the tissue-distribution database that we constructed. The genes exhibiting higher expression in OA cartilage than in normal cartilage and other tissues (a representative EST is shown), or genes exhibiting higher expression in normal cartilage than in OA cartilage and other tissues were selected for further study. Both of these categories of genes were defined as disease and cartilage specific genes.

The disease cartilage specific genes exhibit 200% expression as compared to other tissues. These genes were curated into two groups: Genes that were expressed in all normal or OA pools and 200% (or 50%) as compared to (a) 12-14 other tissues, and (b) 9-11 other tissues, respectively. These genes could be targets for pharmacological intervention or markers.

association studies and positional cloning analysis utilizing microsatellite marker loci, sibling pairs and specific disease-susceptible families. Recently, a large cohort genotypic study suggested that the susceptible gene(s) for the late-onset form of human osteoarthritis has been mapped on chromosome 2, 4, 6, 7, 11 and 16, and the Familial early onset form of the osteoarthritis has been mapped on chromosome 16P between 28 cM and 47 cM, respectively. Rheumatoid arthritis has been mapped to 1(D1S235), 4(D4S1647), 12(D12S373), 16(D16S403 and D16S401), 17(D17S1301), 10(D10S2327 and D10S201), 14(D14S587 and D14S285), which overlap with other autoimmune and inflammatory diseases (Brandi *et al.*, 2001; Ingvarsson *et al.*, 2001; Jawaheer *et al.*, 2001). The QTL approach described above is powerful for nominating chromosomal regions. However, these regions harbor 500-1,000 genes. Breeding strategies in mice have been devised to address this problem and reduce the size of the QTL where the gene identification is more feasible, such as systemic lupus erythematosus (SLE) and arthritis. For example, the Nba2 locus is a major contributor to disease susceptibility in the (NZB x NZW) F1 mouse model of SLE. Kotzin and co-workers generated C57BL/6 mice congenic for this NZB locus, which developed autoantibodies and severe lupus nephritis (Rozzo *et al.*, 2001). Differential gene expression profiling between congenic versus control mice identified IFN-inducible gene (IFI202 and IFI203) within the Nba2 locus (Rozzo *et al.*, 2001).

Another classical example for an inflammatory disease is asthma. A panel of yeast artificial chromosome (YAC) transgenics carrying an asthma QTL was mapped, which reduced the QTL region to one containing only five genes (Symula *et al.*, 1999).

Data Validation

Previous reviews have summarized methods to generate hypothesis driven correlative data and validating these against causative data sets (Attur *et al.*, 2002b and e). However, the ultimate biological validation for therapy comes with the successful phase 3 clinical trials.

Validation procedures for gene mining studies, especially gene expression array data is an essential component for most experimental protocols because of the following difficulties: [1] much of these gene mining technologies have not been compared in parallel with one another using the same clinical samples, [2] recent reports have suggested that some of the commercially available gene expression array have serious flaws in probe design and reproducibility (Knight, 2001), [3] low abundant transcripts (*e.g.* membrane proteins and some cytokines) are not detected in clinical samples when using these gene expression arrays, and [4] variation in the read outs (³²p or fluorescent labeled probes, pseudo-colors or semi-quantitative methods) impede cross-sectional statistical analysis and data integration for data validation. There are some reports which address these issues (Eisen *et al.*, 1998; Rivera *et al.*, 1998; Bassett *et al.*, 1999; Hastie *et al.*, 2000).

Real Time PCR has been a method of choice for validation of mRNA expression. Others and we have

successfully identified and validated low abundant transcripts (5-100 copies of mRNA) in clinical samples that are in the gray zone in gene chip arrays, but were functionally relevant in functional genomic assays (Chudin *et al.*, 2002; Attur *et al.*, 2000b,d). However, the limitation of this technology is encountered for functionally active molecules with transcripts ≤ 5 per cell (Mahadevappa and Warrington, 1999). Amplification with RT-PCR using more than 30 cycles is useful. Sub cellular localization of differentially expressed genes by *in situ* staining using antibodies or riboprobes is essential. This has recently been demonstrated using differentially expressed genes/proteins such as Osteopontin, Erg-1, MMP-1, 3, 8, 9 and 13 in arthritis (Wang *et al.*, 2000; Tetlow *et al.*, 2001). Gene expression patterns of MMP-1, 3, 9 and aggrecanase in OA-affected chondrocytes in cartilage were found to be zonal and grade specific (Freemont *et al.*, 1997). These approaches can allow validation of transcripts/proteins across an array of sections of clinical samples.

Complex diseases have a tendency to show variability in disease-associated genes among populations (Strohan, 2002). This problem can be rectified by pooling samples (Figure 2). In general, 2% of the transcripts were found to be upregulated and 3% downregulated in comparisons between normal (n= 20) and OA-affected cartilage (n=70). Among these, approximately 20% represented receptors, transcription factors and enzymes, which may have therapeutic implications. Furthermore, these differentially expressed transcripts (TGF β , IL-8, IL-6 and TACE) can again be validated in individual sample of disease (Attur *et al.*, 2002a). The differentially expressed transcripts were classified into levels 1 to 5, representing all or none expression in control and disease tissue to 60% representation. We have identified various genes in these different categories. Furthermore, potential targets from these levels can also be evaluated for tissue distribution as shown in Figure 3b. These approaches help to identify and develop diseases and tissue (cartilage) specific targets for pharmacological intervention or markers.

Functional Genomics

Functional genomic analysis involves a systematic effort to understand the function of genes and gene products (transcripts and proteins) and biological systems (cell, tissue or organism) classically performed for single genes (*e.g.* generation of mutants, analysis of proteins and transcripts) in the context of the whole genome. Functional genomics can be conceptually divided into two matrix approaches: (a) Gene-driven approach, where one uses genomic information for identifying, cloning, expressing and characterizing the gene at the molecular level, (b) Phenotype-driven approach, which analyzes phenotypes from random mutation screens or naturally occurring variants (mouse mutants, human diseases) to identify and characterize genes for the phenotype, without prior knowledge of the underlying molecular mechanism or function. Both strategies are complimentary leading collectively to association of phenotype with genotypes. As functional genomics begins to mature into a coherent science (as Molecular Biology did in the last half of the

century) its constituent fields become clearer. They include bioinformatics, structural genomics, comparative genomics, expression genomics and proteomics.

The following review focuses on bioinformatic and traditional biological approaches to analyze expression data with emphasis on functional genomics of genes belonging to MMPs, matrix protein/proteoglycans and cytokine and cytokine receptor family in arthritis with respect to cartilage biology.

Analysis of Differentially Expressed Extracellular Matrix Components (ECM) in Cartilage

Cartilage is an aneural, alymphatic and avascular tissue, which constitute chondrocytes and the surrounding ECM. ECMs regulate some of the most fundamental cellular processes such as growth, survival, differentiation, motility, signal transduction and cell shape in cartilage (Attur *et al.*, 2000; Loeser, 2000; Scully *et al.*, 2001). Chondrocytes when grown in culture (outside the cartilage matrix) *in vitro*, dedifferentiates into fibroblast-like cells (Schnabel *et al.*, 2002; Stokes *et al.*, 2002). Various receptor/matrix interactions have been reported in chondrocytes, which maintain a chondrocyte phenotype in cartilage (Heinegard *et al.*, 1989; 1998). Several ECMs reported to be differentially expressed in normal and arthritis-affected cartilage. These include fibronectin, collagens, proteoglycans, cartilage oligomeric matrix protein (COMP), Osteopontin (OPN) and vitronectin (Clark *et al.*, 1999; Pullig *et al.*, 2000; Attur *et al.*, 2001; 2002a; Aigner *et al.*, 2001, Aigner and McKenna, 2002). Examination of the role of matrix proteins/proteoglycans in human chondrocyte and cartilage homeostasis requires an unconventional assay system.

In view of these observations, we have developed an *ex vivo* human cartilage organ culture assay to examine the role of chondrocyte/matrix interaction without disturbing this delicate architecture. These arthritis-affected cartilage samples spontaneously release various inflammatory mediators, including nitric oxide (NO), PGE₂, MMPs, cytokines and demonstrate various dynamics in matrix homeostasis *ex vivo* (Amin *et al.*, 1999a,b; 2000; Abramson *et al.*, 2001). Recombinant proteins such as ligands, soluble receptors, antibodies, and other low molecular weight disease modifying drugs (DMARDs) can be added to this assay to analyze inflammation and cartilage homeostasis. This assay has recently been extended into pharmacological and chemogenomic assays for profiling transcripts in the presence of NSAIDs, DMARDs and lead drug candidates, which have the ability to modify cartilage homeostasis (Amin *et al.*, 1999b; Amin, 2000). This assay not only facilitates understanding functional genomics of potential targets but also helps to validate the data in the human system as described below.

Functional Analysis of Fibronectin and Osteopontin in Cartilage

Fibronectin (FN) and osteopontin (OPN) are differentially expressed in normal and arthritis-affected cartilage (Pullig *et al.*, 2000; Attur *et al.*, 2001). FN and OPN were identified

as genes on a two different chromosomes 2q and 4q associated with nodal OA and SLE respectively. (Wright *et al.*, 1996; Forton *et al.*, 2002). Furthermore, fragments of FN protein and OPN have been reported to act as pro-inflammatory and anti-inflammatory mediators respectively in human cartilage (Saito *et al.*, 1999; Attur *et al.*, 2001; 2000c).

The integrin receptors for FN ($\alpha 5\beta 1$) and OPN ($\alpha v\beta 3$) respectively have been identified in chondrocytes. Binding of monoclonal antibody, (which acts as agonist similar to FN-N-terminal fragment) to $\alpha 5\beta 1$, upregulates the inflammatory mediators as well as the cytokines. In contrast, an antibody to $\alpha v\beta 3$, which acts as an agonist similar to OPN, attenuates the production of IL-1 β (triggered by $\alpha 5\beta 1$, IL-1 β and IL-18) in a dominant negative fashion in cartilage. These data demonstrate a cross talk in signaling mechanisms among integrins mediated via cartilage matrix components. It is interesting to note that $\beta 1$ integrin null mice showed diminished cartilage development (Ekholm *et al.*, 2002)

These regulatory circuits demonstrate the pivotal role of chondrocyte receptor/matrix interaction. These dysfunctional signaling mechanisms influence cartilage homeostasis and a provocative role in the pathogenesis of osteoarthritis (Attur *et al.*, 2000c; Denhardt *et al.*, 2001).

Functional Genomic Studies of Matrix Proteins by a Phenotype-driven Approach

Synovial hyperplasia has been observed predominantly in RA rather than OA, despite differences in underlying etiologies of the two disorders. The autosomal recessive disorder camptodactyly arthropathy-coxa vara-pericarditis (CACP) affects the joints and shows synoviocyte hyperplasia. Using a positional-candidate gene approach, Marcelino *et al.* identified mutations in the human gene encoding a secreted proteoglycan previously identified as both "megakaryocyte-stimulating factor precursor" and "superficial zone protein" in individuals affected with CACP (Marcelino *et al.*, 1999). These proteins contain domains that have homology to somatomedin B, heparin-binding proteins, mucins and haemopexins. This CACP protein may be involved in regulating cell cycle and growth. Its dysfunctional expression may be involved in hyperplasia of synovium, pericardium and pleura as observed in arthritis. "CACP knock out" mouse shows similar cartilage destruction and synovial hyperplasia as patients with CACP with no infiltrating inflammatory cells (unpublished data and personal communication by Jose Marcelino). This gene product may be a potential target in arthritis.

Similarly, Hurvitz *et al.* have also mapped the WISP3 gene (using a positional candidate approach) for progressive pseudo-rheumatoid dysplasia (PPD), which was previously misdiagnosed as juvenile RA (Hurvitz *et al.*, 1999). PPD, like arthritis, shows loss of normal cell columnar organization in the growth zone in the subchondral region of cartilage. WISP3 genes are members of the connective tissue growth factor family, which are secreted and matrix-bound. At least nine different mutations were identified in WISP3 affected individuals. The normal function of WISP3 is unknown. Animal "knock

out" or transgenic studies may facilitate in understanding the role of WISP3 in abnormal conditions. Furthermore, more refined technologies of advanced conditional knock in/out using Lac I repressor will allow targeting of endogenous loci, switching them on and off repeatedly to create reversible models of human diseases and normal development in the mouse (Cronin *et al.*, 2001).

Characterization of Developmental Genes in Cartilage and Bone

The chick embryo and zebra fish are two models that are extensively utilized to characterize developmental genes in bone and cartilage (Cancedda *et al.*, 2000; Kimmel *et al.*, 2001). Ito *et al.*, using subtractive hybridization, have recently cloned a cDNA coding for normal lysyl oxidase related protein named LOXC in differentiated and calcified cells (Ito *et al.*, 2001). The deduced amino acid sequence of LOXC contained 50% identity to the Mouse lysyl oxidase. The expression of LOXC mRNA and protein levels increased in the hypotrophic and calcified chondrocytes in the growth plate in adult mice. Transduction of the full length LOXC cDNA resulted in expression of lysyl oxidase activity in both type I and type II collagen derived chick embryos, which could be inhibited by β -amino propionitrile, a specific inhibitor of lysyl oxidase. These data suggest that LOXC possesses lysyl oxidase enzymatic activity, which may be involved in the cross-linking of the extracellular matrix. However, the possible role of LOXC in endochondral bone formation cannot be ruled out. Similarly, the overexpression of c-myc oncogene increase cell size and impairs cartilage differentiation during chick limb development (Piedra *et al.*, 2002).

The Zebra Fish

Zebra fish serves as a powerful experimental model for the genetic dissection of genes for functional genomics as illustrated by recent large-scale ENU-mutagenesis study resulting in identification of developmentally important genes (Childs *et al.*, 2002). In zebra fish, the cartilages of the pharynx develop during late embryogenesis and grow extensively in the larvae before eventually being replaced by bone. One can examine chondrocyte arrangement, shape, number and division in cartilage in this system (Kimmel *et al.*, 1998).

Several technologies have been developed to manipulate genes (knock-in/knock-out) to examine the phenotypes in a relatively short life span of the fish (Oates *et al.*, 1999). One such example is expression profile for chondromodulin-1, which could be followed in the late developmental stages cartilage and chondrogenic region of the pectoral fin (Sachdev *et al.*, 2001). Additionally, disruption of endothelin disrupts development of the lower jaw and other ventral cartilage in pharyngeal segments (Miller *et al.*, 2000). Injection of retinoic acid disrupts craniofacial morphogenesis in zebra fish (Yan *et al.*, 1998) and exposure to dioxin (TCDD) disrupts cartilage growth (Teraoka *et al.*, 2002).

In summary, the complete zebra fish genome will be sequenced before the end of this year and will serve as a

useful model to study function of novel genes in cartilage development and homeostasis.

Functional Genomic Analysis of Mutated Collagens in Murine Models

Chondrocytes express collagen type I, II, III, V, VI, IX, X, XII and XIV depending on their physiological stage (Petit *et al.*, 1998). Approximately 278 different mutations have been reported to date in genes for type I, II, III, IX, X and XI collagens from unrelated arthritis-affected patients. A majority (78%) of the mutations are single-base and either change the codon of a critical amino acid or lead to abnormal RNA splicing, which may lead to a spectrum of diseases of bone and cartilage including osteogenesis imperfecta, a variety of chondrodysplasia and OA (Kuivaniemi *et al.*, 1997).

The laboratory mouse is a powerful and wide-ranging genetic tool, which can be utilized as a major experimental model for studying mammalian gene functions *in vivo* and modeling human disease traits. Collagen type IX, non-fibrillar collagen localized on the surface of type II collagen is well studied in mouse model. Two alternate spliced forms of collagen type IX are expressed on hyaline cartilage. A mouse strain lacking both forms showed no detectable abnormalities at birth, but develop a severe non-inflammatory degenerative joint disease resembling human OA (Fassler *et al.*, 1994). Independent experiments by Nakata *and coworkers* using a tissue-specific promoter/enhancer to express type IX collagen also revealed pathological changes similar to OA and chondrodysplasia as observed in humans (Nakata *et al.*, 1993). These studies demonstrate the complex interactions of matrix components in a disease process.

Identification of Novel Proteases from Human Arthritis-Affected Cartilage

The analysis of the human genome has allowed us to predict the presence of ~ 500 MMP-like transcripts in humans, which need to be characterized, with respect to their function. Human arthritis-affected cartilage and synovium are one of the richest sources of differentially expressed disease-specific proteases (Freemont *et al.*, 1997; Patel *et al.*, 1998; Tetlow *et al.*, 2001). Gene mining efforts using total RNA from normal, rheumatoid and osteoarthritis-affected cartilage (Patel *et al.*, 1998), yielded a clone (clone 8), for a protein which showed a partial cysteine switch sequence (PKVGY) and zinc binding region (HELGHN) separated by 690 base pairs similar to that seen in snake venom proteases (Wolfsberg *et al.*, 1993). This was classified as a unique protease because preliminary characterization of this protease excluded it from the matrixin family of MMPs.

A bioinformatic approach was utilized to identify the protease with a hypothesis in view that the structure/function of a protein domain shows evolutionary conservation and, by convention, is represented by a distinct geometric shape. A library of curated protein domains with their biological descriptions is available through the Pfam and SMART databases (Sonnhammer

et al., 1997; Schultz *et al.*, 1998). Using the above concept and databases, comparative genomics and bioinformatic approaches were further combined to compare the 3-D ribbon structures of clone 8, in spite of its low sequence homology with other proteases in the database. The three hydrophobic side chains that support 3-D folds were conserved in snake venom protease and clone 8, suggesting that it was structurally and functionally similar to M12B snake venom protease family. The full-length cDNA sequence of this cartilage snake venom protease showed ~ 99% homology to TNF α convertase (TACE) (Patel *et al.*, 1998). The specificity of TACE was confirmed by its ability to cleave membrane bound proTNF α from soluble TNF α . Other putative substrates for TACE include L-selectin, TNF α receptor I and II, APP, IL-1 receptor and IL-6 receptor (Moss *et al.*, 2001). How does TACE distinguish between all of these substrates? Preliminary data suggests that different domains of TACE are necessary for turnover of different substrates (Moss *et al.*, 2001). Inhibitors of TACE block TNF α activity in arthritis-affected cartilage. These experiments demonstrated a functional paracrine/autocrine role of TNF α in arthritis-affected cartilage that may depend, in part, on upregulated levels of chondrocyte-derived TACE (Patel *et al.*, 1998). TACE is a potential target for pharmacological intervention of TNF α production and therefore arthritis (Newton *et al.*, 2002; Attur *et al.*, 2002c).

Serine Proteases

Proteins are made up of one or more building blocks or "domains", depending on the number or types of the domains, proteins exhibit different biological capabilities. Conserved serine proteases are common denominators in various proteins exhibiting various biological activities. These proteins include plasminogen, apolipoprotein (A), urokinase-type plasminogen activator, prostate-specific antigen, coagulation factor XI, coagulation factor X and complement C1r component. These molecules share a common denominator with respect to their ability to have serine protease activity, but show domain shuffling due to other heterogenous domains. These are also called plasma proteases of coagulation and complement systems. The ancient trypsin family serine protease domain occurs in combination with a myriad of protein interaction domains. Most of these domains are evolutionarily ancient, that is, with the exception of the Gla domain (Subramanian *et al.*, 2001). Other serine proteases have been identified by differential expression of transcripts in human normal and arthritic tissues *e.g.* human High-temperature requirement A (HtrA), an evolutionarily conserved serine protease (Hu *et al.*, 1998). Cloning and expression of human HtrA exhibited endoproteolytic activity, including autocatalytic cleavage. The putative active site was mapped to serine 328. Recent substrate specificity studies suggest that this protease has the ability to degrade COMP and fibronectin, which are major structural proteins found in human cartilage (Ganu *et al.*, 2001).

Identifying the Role of Matrix Metalloproteases (MMPs) by a Phenotype-driven Functional Genomic Approach

Familial osteolysis is a rare inherited disorder where affected individuals exhibit characteristic facial features, lytic lesions of the bone and arthritis. Linkage analysis showed MMP-2 as a candidate gene harboring mutation in four Saudi Arabian families. However, MMP-2-null mice have no developmental defect, but mice targeted MT1-MMP gene show the same features as individuals with osteolysis and arthritis (Zhou *et al.*, 2000; Martignetti *et al.*, 2001). One explanation of the difference between man and rodent may be due to distinct regulation of MMP-2 and MT1-MMP, and the differential balanced regulations they exert on latent TGF β , which regulates cartilage degradation. This example highlights the limitation of mouse gene targeting methods despite their exquisite potential for addressing gene functions in defined contexts.

Functional Genomics by Transgenic Reporter Mice

Type II collagen has been one of the genes identified to be dysfunctional in human OA. It may be influenced by environmental factors and epistasis. Cho *et al.* 2001, have developed a Col-2- GFP reporter mouse as a new tool to study cartilage and skeletal development. The cartilage and bone biology can be assessed throughout the body including non-skeletal cartilaginous structure such as external ears. This model also allows one to evaluate the role of chondrocytes in synthesizing templates for skeletal development and chondrogenesis in real time and offers the potential to monitor dynamic events during at least short periods during pharmacological intervention and other environmental conditions.

Analysis of Cytokines and Their Receptors in Arthritis

In view of the importance of IL-1 in OA and RA, several homologs of IL-1 and its receptors have been identified from the databank using electronic mining. These include ST2, an IL-1 receptor homolog, and IL-1H1-H4 (Mulero *et al.*, 1999; Kumar *et al.*, 2000; Lin *et al.*, 2001). Gene array analysis of human normal and arthritis-affected cartilage showed mRNA expression of IL-1 receptor accessory protein (IL-1RAcP) and IL-1 type I receptor (IL-1RI), but not IL-1 antagonist (IL-1ra) and IL-1 type II decoy receptor (IL-1RII). Similarly, human synovial and epithelial cells also showed low expression of IL-1RII mRNA (Attur *et al.*, 2000b).

Gene Therapy Approach for Functional Genomics *in vitro*

Low amounts (pg/g cartilage) of IL-1, which is released in human OA-affected cartilage (Attur *et al.*, 2000b), during early stages of the disease, have the ability to act unopposed with respect to the lack of naturally occurring IL-1 antagonistic activity in the cartilage and inflict detrimental effects on cartilage homeostasis in long-term diseases such as osteoarthritis.

Functional analysis showed that recombinant soluble

(s) IL-1RII, but not soluble TNF receptor:Fc, significantly inhibited IL-1 β -induced inflammatory mediators in chondrocytes, synovial and epithelial cells. Reconstitution of human IL-1RII expression in various IL-1RII-deficient cell types by adenovirus expressing human IL-1RII showed expression of membrane IL-1RII (mIL-1RII) and spontaneous release of functional soluble IL-1RII (sIL-1RII) and rendered the IL-1RII⁺ cells resistant to autocrine and exogenous IL-1 induced inflammatory mediators or decrease in proteoglycan synthesis. In co-cultures, IL-1RII⁺ synovial cells released a functional sIL-1RII, which in a paracrine fashion protected chondrocytes from the effects of IL-1. Furthermore, autologous IL-1RII⁺ (but not IL-1RII⁻) chondrocytes when transplanted onto human OA-cartilage *in vitro* [showed spontaneous release of sIL-1RII for 20 days], and inhibited the spontaneous production of inflammatory mediators in cartilage in *ex vivo* conditions.

In summary, reconstitution of IL-1RII in IL-1RII⁻ cells using gene therapy approaches, significantly protects IL-1RII⁻ cells against the autocrine/paracrine effects of IL-1 β by acting at several levels of IL-1 signaling and transcription (Attur *et al.*, 2000b).

A Gene Therapy Approach to Functional Genomics *in vivo*

Polymorphism in the IL-1 β gene is associated with inflammation in arthritis (Moos *et al.*, 2000), TGF β 1 gene is associated with spinal OA (osteophytosis) and IGF-1 gene is associated with generalized OA (Meulenbelt *et al.*, 1998). Differential expression of mRNA in normal and arthritis-affected cartilage also showed modulation of IL-1 β , TGF β 1 and IGF-1 mRNA transcripts (Meulenbelt *et al.*, 1998; Yamada *et al.*, 2000). The role of IL-1, TGF β and IGF in joints can be assessed using a gene therapy approach in a collagen induced arthritis model and rabbit models.

Constitutive intra-articular expression of an adenoviral IL-1 transgene in rabbit joints induces multiple intra-articular manifestations, which include intense inflammation, leukocytosis, synovial hypertrophy, hyperplasia, highly aggressive pannus formation, erosion of cartilage and bone. It also induced systemic effects including diarrhea and fever. Following the loss of the transgene, (which occurs after 28 days) most of the pathophysiological symptoms described above subsided within 4 weeks (Ghivizzani *et al.*, 1997).

In spite of some of its beneficial effects on chondrocyte metabolism, over-expression of TGF β by adenovirus in rodent joints showed the formation of osteophytes (Van den Berg, 1995) and deregulation of bone remodeling (Smith *et al.*, 2000). Gene transfer of IGF-1 into rabbit knee joints promotes proteoglycan synthesis without significantly affecting inflammation or cartilage breakdown. This local gene transfer of IGF-1 to joints could serve as a therapeutic strategy to stimulate new matrix synthesis in both RA and OA (Mi *et al.*, 2000). Other target genes (TSG-6, IL-4, SOD, IL-1RA, IL-1RII and p16^{INK4a}) have also been tested in this model system with success (Taniguchi *et al.*, 1999; Bardos *et al.*, 2001; Iyama *et al.*, 2001; Woods *et al.*, 2001). In summary, a gene therapy approach (*in vitro* or *in vivo*)

allows identification of the function of candidate genes identified from a genomic screen in a complex cartilage and joint environment.

An In Vivo Model of SCID Mouse for Human Synovium/ Cartilage Invasion for Functional Genomics in RA

Synovial hypertrophy and pannus formation play a critical role in inflammation and cartilage destruction in RA and OA. In view of this, S. Gay and colleagues have developed a human synovial fibroblast/cartilage interaction SCID model *in vivo* to examine the role of various genes (Jorgensen and Gay, 1998). Briefly, human RA-affected synovial fibroblasts are grown *in vitro* and transfected with a gene of choice. The cells are then packaged in an inert sponge together with normal human cartilage and implanted in a renal capsule in mice. This strategy has been utilized to examine the effect of IL-1Ra, p55-TNF α receptor, IL-10, tumor suppressors PTEN and p53 in cartilage degradation (Jorgensen and Gay, 1998; Attur *et al.*, 2000a).

Angiogenesis in Rheumatoid Arthritis (RA)

The role of angiogenesis in OA and RA-affected synovium has recently been reviewed (Koch *et al.*, 1998). These pro-angiogenic factors have become targets for treatment of RA, a strategy which has been shown to work efficiently in animal models (Scola *et al.*, 2001). LM609 (an antagonist antibody to α v β 3) blocked angiogenesis in human breast cancer and synovial hypertrophy in rabbit models of RA (Brooks *et al.*, 1995; Storgard *et al.*, 1999).

Plasmin is essential for MMP activation, endothelial cell migration and degradation of extracellular matrix. The process is also common to neoangiogenesis pannus formation and cartilage degradation in the joint. A gene therapy approach was utilized to examine a hypothesis based on these observations. Adenovirus-mediated gene transfer of urokinase plasminogen inhibitor reversed angiogenesis in experimental arthritis (Apparailly *et al.*, 2002). Several proinflammatory/neoangiogenic factors such as PGE2, NO and VEGF have been reported to be involved in synovial hypertrophy. Similarly, angiopoietins (Ang-1 and -2), ligands which stabilize vascularization during angiogenesis, have been reported to be expressed in RA synovium. These factors may also be potential targets for RA therapy (Scott *et al.*, 2002).

Progressing Beyond Single Genes: Environmental Impact and Epistasis

Genes operate in environments. These environments can range from cellular location, to the specific forms (alleles) of other genes expressed elsewhere in the genome, to the characteristics of the room in which a behavior is assessed. RA is multi-factorial disease determined by both genetic and environmental factors. Recently a DANISH nationwide study conducted in twin population suggest that environmental effects may be more important than genetic effects which may crossover with other autoimmune diseases. (Jawaheer *et al.*, 2002; Svendsen *et al.*, 2002).

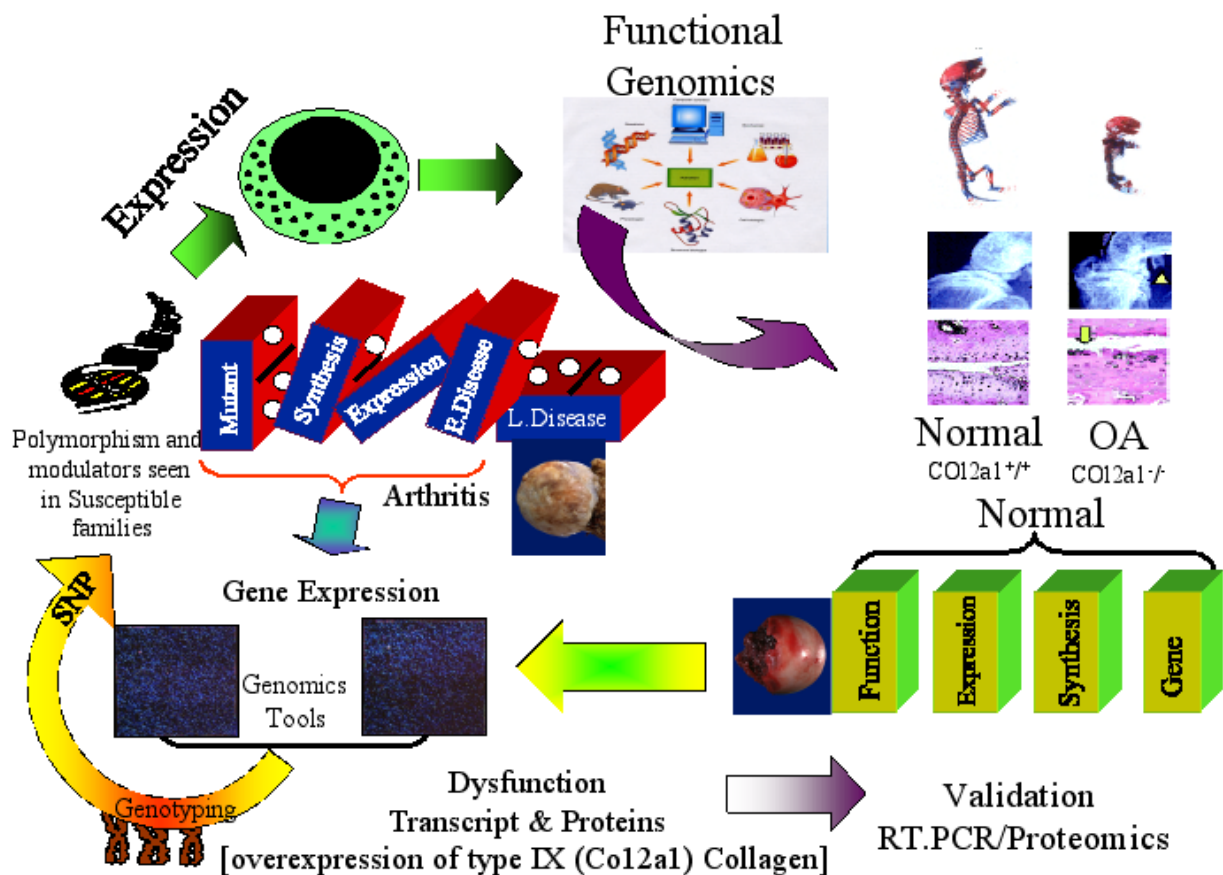


Figure 4. Hypothetical scheme showing hybrid approach to functional genomics in OA. Normal expression of genes in cartilage shows formation of normal hyaline cartilage as observed. However, overexpression of a dysfunctional gene (e.g. Type IX (Col2a1) collagen) initiates a domino-effect in long-term diseases when the phenotype is observed in the later stage of life and leads to destruction of cartilage as shown. Gene expression array of normal and OA-affected cartilage identifies such dysfunctional expression of such transcripts, which can be validated with Real Time PCR and proteomics using a separate set of normal and OA-affected cartilage samples (Attur, unpublished data). Genotyping and/or SNP analysis also identified mutations in susceptible families of the same gene (Barat-Houari *et al.*, 2002; Doris, 2002; Schmidt *et al.*, 2002). The dysfunctional expression can be mimicked in cells by knock in/out technologies and finally validated in mice using tissue specific modulation of the gene (van Meurs *et al.*, 1999; Barton *et al.*, 2002). This strategy exemplifies a hybrid approach to genomics using complementary technologies and approaches.

The genetic effects (epistasis) in arthritis have been observed in animal models. The effects of genetic manipulations such as targeted gene deletions and transgenic overexpression of genes can vary widely depending upon the genetic make-up of the animal carrying the targeted gene. For example, the interaction between different forms of collagens in matrix plays an important role in cartilage homeostasis. Although mice deficient in collagen II (Col2a1^{-/-}) die at birth, and Col9a1^{-/-} mice develop OA-like phenotypes, the Col2a1^{+/-}Col9a1^{-/-} mice show no accelerated OA (Aszodi *et al.*, 2000; 2001). These observations suggest that extracellular matrix proteins have different roles influenced by epistasis during cartilage development.

Similarly, in FcγRIIB knock out mice are susceptible to develop SLE in C57BL/6 mice due to a linked sle1 locus. B6.RIIB^{-/-}/lpr mice are protected from disease progression despite high titer of anti-nuclear antibodies (ANA). In contrast, B6.RIIB^{-/-}/yaa mice has significantly enhanced

disease despite reduced ANA. These study identified two novel recessive loci required for ANA phenotype, which indicate the epistatic property of this SLE model (Bolland *et al.*, 2002). Thus, development and progression of arthritis is dependent on both the environment and genetic factors.

Conclusions and Future Directions

“System Biology” or “Whole-istic Biology” is a concept that has pervaded all fields of science and penetrated into popular thinking. It is not a new concept. Ludwig van Bertalanffy proposed “General System Theory” in psychology, economics and social sciences back in 1940. The post-genomic revolution has redefined the concept. Rightly so, successful analysis of complex human diseases such as arthritis will require understanding of the functional interactions between key components of cells (such as chondrocytes and synovial cells), organs (synovium and cartilage) and systems (mobile joint) as well as the

interactions that change in the disease state (clinical material and diagnosis). This information resides neither in the genome or individual gene(s)/protein(s), but it seems to lie at the level of protein interactions within the context of subcellular, cellular, tissue, organ and system structure (summarized in Figure 1).

Thus, to identify novel targets for pharmacological intervention, diagnosis and prognosis, a simple association of gene expression with disease (*e.g.* generated through a gene array) does not validate the gene(s) as a target(s) in the disease. Even a human genetic approach to identify targets associated with the disease does not necessarily generate chemically tractable molecular targets. Rather the goal of target validation and functional genomics is to strengthen correlative data (from gene arrays, EST libraries and proteomics) by demonstrating a causal role for the candidate in a disease model. Bioinformatics is an important adhesive tool in functional genomics that will help bridge the gap between correlative data and causative data although there are limitations in predicting *ab initio* gene structure, gene function and protein folds from the raw sequence data. Clearly, a lot needs to be done, as more than 40% of the 35,000 genes (and possibly 120,000 different proteins they may code) have not been ascribed any functional attribute (Yaspo, 2001), either a biochemical function (*e.g.* kinase), a cellular function (*e.g.* a specific signaling pathway) or a function at the tissue/organism level (*e.g.* brain development, immune response, etc.).

Successful drug treatments of the past and present involve fewer than 500 targets including growth factors and cytokines as of 1996 (Bumol and Watanabe, 2001). It is assumed that at least 5000 of the possible 120,000 proteins may be potential therapeutic proteins or targets, suggesting that only 10% of potential therapeutic strategies have been identified and exploited to date (Drews, 2000). The gene(s) involved in the etiology of arthritis, subcategorized based on their clinical symptoms, still remain to be identified and characterized among these targets. This is a formidable task as at least 14 different linkages to high-density markers on different chromosomes have been identified for hand, hip and knee OA (Bateman, 2002). Collagen IX and XI remain tantalizing candidate genes in OA risk. Gene expression data between normal and OA-affected cartilage show an "inflammatory/proliferative" dysfunctional gene signature comprising of over 1,500 transcripts (unpublished data). Such gene mining efforts with bioinformatics will facilitate co-relating gene expression data with clinical outcomes as described in normal and RA-affected monocytes and cartilage (Stuhlmuller *et al.*, 2000). These preliminary studies may lead to predictive medicine in arthritis, identifying different disease states (*e.g.* therapy-induced remission) with respect to modulation of novel genes.

Arthritis is now a disease that is challenged with many drugs. On the whole, these drugs treat inflammation and pain as a symptom, but do not address the actual cause of the disease. Some of the new generation of drugs, which also target symptoms of the disease, including vascular adhesion protein 1 (VAP-1) and vascular endothelial growth factor (VEGF) and its receptor Flt-1 (Gerber *et al.*, 1999).. Gene therapy approaches in human RA have given some

promising results in Phase II clinical trials (Evans *et al.*, 2001) Increasing the understanding of molecular cascades involved in the disease processes by genomic approaches will allow us to produce significantly better drugs than in the past with increased selectivity and fewer side effects.

In summary, the convergent evolution of subcategories of genomic analysis such as development of computing hardware, algorithms and databases, have made it possible to explore functionality in a quantitative manner all the way from the level of the gene to the cell to the physiological functions of whole organs and regulatory systems (Davidson *et al.*, 2002; Kitano, 2002; Noble, 2002). Genomics in this century is thus posing to be a highly quantitative and computer-intensive discipline.

Acknowledgements

We thank Andrea L. Barret for preparation of this manuscript, Sonali Trivedi for preparation of the figures, and Dr. Smita Palejwala for critically reviewing the manuscript.

References

- Abramson, S. B., Attur, M., Amin, A. R., and Clancy, R. 2001. Nitric oxide and inflammatory mediators in the perpetuation of osteoarthritis. *Curr. Rheumatol. Rep.* 3: 535-541.
- Agrawal, D., Chen, T., Irby, R., Quackenbush, J., Chambers, A. F., Szabo, M., Cantor, A., Coppola, D., and Yeatman, T. J. 2002. Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. *J. Natl. Cancer Inst.* 94: 513-521.
- Aigner, T., Zien, A., Gehrsitz, A., Gebhard, P. M., and McKenna, L. 2001. Anabolic and catabolic gene expression pattern analysis in normal versus osteoarthritic cartilage using complementary DNA-array technology. *Arthritis Rheum.* 44: 2777-2789.
- Aigner, T. and McKenna, L. 2002. Molecular pathology and pathobiology of osteoarthritic cartilage. *Cell Mol. Life Sci.* 59: 5-18.
- Amin, A. R., Attur, M., and Abramson, S. B. 1999a. Nitric oxide synthase and cyclooxygenases: distribution, regulation, and intervention in arthritis. *Curr. Opin. Rheumatol.* 11: 202-209.
- Amin, A. R., Attur, M. G., and Abramson, S. B. 1999b. Regulation of nitric oxide and inflammatory mediators in human osteoarthritis-affected cartilage: implication for pharmacological intervention. In GM Rubanyi (ed), ed., *The Pathophysiology and Clinical Applications of Nitric Oxide.* pp. 397-413. Harwood Academic Publishers, Richmond, CA.
- Amin, A. R., Dave, M., Attur, M., and Abramson, S. B. 2000. COX-2, NO, and cartilage damage and repair. *Curr. Rheumatol. Rep.* 2: 447-453.
- Amin, A.R. 2000. Gene Mining, Bioinformatics and Functional Genomics in Human Arthritis and Inflammatory Diseases *Ex Vivo.* *Drug Develop. Res.* 49: 22-28.
- Apparailly, F., Bouquet, C., Millet, V., Noel, D., Jacquet, C., Opolon, P., Perricaudet, M., Sany, J., Yeh, P., and Jorgensen, C. 2002. Adenovirus-mediated gene transfer

- of urokinase plasminogen inhibitor inhibits angiogenesis in experimental arthritis. *Gene Ther.* 9: 192-200.
- Aszodi, A., Hunziker, E. B., Olsen, B. R., and Fassler, R. 2001. The role of collagen II and cartilage fibril-associated molecules in skeletal development. *Osteoarthritis and Cartilage.* 9 Suppl A: S150-S159.
- Aszodi, A., Bateman, J. F., Gustafsson, E., Boot-Handford, R., and Fassler, R. 2000. Mammalian skeletogenesis and extracellular matrix: what can we learn from knockout mice? *Cell Struct. Funct.* 25: 73-84.
- Attur, M. G., Patel, I. R., Patel, R. N., Abramson, S. B., and Amin, A. R. 1998. Autocrine production of IL-1 β by human osteoarthritis-affected cartilage and differential regulation of endogenous nitric oxide, IL-6, Prostaglandin E₂ and IL-8. *Proc. Assoc. Amer. Physicians.* 110: 1-8.
- Attur, M. G., Bingham, C. O. I., Dave, M. N., Abramson, S. B., and Amin, A. R. 2000a. Model Protocol to Study Pharmacogenomics in Inflammatory Diseases: Human Rheumatoid Arthritis. *Drug Development Research* 49: 29-33.
- Attur, M. G., Dave, M., Cipolletta, C., Kang, P., Goldring, M. B., Patel, I. R., Abramson, S. B., and Amin, A. R. 2000b. Reversal of autocrine and paracrine effects of interleukin 1 (IL-1) in human arthritis by type II IL-1 decoy receptor. Potential for pharmacological intervention. *J. Biol. Chem.* 275: 40307-40315.
- Attur, M. G., Dave, M. N., Clancy, R. M., Patel, I. R., Abramson, S. B., and Amin, A. R. 2000c. Functional genomic analysis in arthritis-affected cartilage: yin-yang regulation of inflammatory mediators by $\alpha 5\beta 1$ and $\alpha V\beta 3$ integrins. *J. Immunol.* 164: 2684-2691.
- Attur, M. G., Dave, M. N., Stuchin, S., Kowalski, A. J., Steiner, G., Abramson, S. B., Denhardt, D. T., and Amin, A. R. 2001. Osteopontin: an intrinsic inhibitor of inflammation in cartilage. *Arthritis Rheum.* 44: 578-584.
- Attur, M. G., Dave, M., Akamatsu, M., Katoh, M., and Amin, A. R. 2002a. Osteoarthritis or osteoarthrosis: the definition of inflammation becomes a semantic issue in the genomic era of molecular medicine. *Osteoarthritis and Cartilage.* 10: 1-4.
- Attur, M.G., Dave, M.N., and Amin, A.R. 2002b. Genomining and Functional Genomic in Human Osteoarthritis. *Curr. Genomics.* In Press.
- Attur, M. G., Dave, M. N., Patel, I. R., Abramson, S. B., and Amin, A. R. 2002c. Regulation of inflammatory mediators by tetracyclines. In *Tetracyclines in Biology, Chemistry and Medicine* p. 293.
- Attur, M. G., Dave, M. N., Leung, M. Y., Cipolletta, C., Meseck, M., Woo, S. L. C., and Amin, A. R. 2002d. Functional genomic analysis of type II IL-1 β decoy receptor: Potential for gene therapy in human arthritis and inflammation. *J. Immunol.* 168: 2001-2010.
- Attur, M., Dave, M., and Amin, A. R. Functional Genomics in Arthritis in the Post-Genomic Era of Medicine. *J. Pharmacogenomics.* 2002e. In Press
- Bakker, N. P., van, E. M., Zurcher, C., Faaber, P., Lemmens, A., Hazenberg, M., Bontrop, R. E., and Jonker, M. 1990. Experimental immune mediated arthritis in rhesus monkeys. A model for human rheumatoid arthritis? *Rheumatol. Int* 10: 21-29.
- Barat-Houari, M., Clement, K., Vatin, V., Dina, C., Bonhomme, G., Vasseur, F., Guy-Grand, B., and Froguel, P. 2002. Positional candidate gene analysis of lim domain homeobox gene (*isl-1*) on chromosome 5q11-q13 in a French morbidly obese population suggests indication for association with type 2 diabetes. *Diabetes* 51: 1640-1643.
- Bardos, T., Kamath, R. V., Mikecz, K., and Glant, T. T. 2001. Anti-inflammatory and chondroprotective effect of TSG-6 (tumor necrosis factor-alpha-stimulated gene-6) in murine models of experimental arthritis. *Am. J. Pathol.* 159: 1711-1721.
- Barton, E. R., Morris, L., Musaro, A., Rosenthal, N., and Sweeney, H. L. 2002. Muscle-specific expression of insulin-like growth factor I counters muscle decline in *mdx* mice. *J. Cell Biol.* 157: 137-148.
- Bassett, D. E. J., Eisen, M. B., and Boguski, M. S. 1999. Gene expression informatics—it's all in your mine. *Nat. Genet.* 21: 51-55.
- Bateman, 2002. *New Trends in Osteoarthritis.* Eds. Carrabba, M. and Puttini, P. S. pp: 55 – 59. Proceedings of the International Congress. Milan (Italy).
- Bolland, S., Yim, Y. S., Tus, K., Wakeland, E. K., and Ravetch, J. V. 2002. Genetic Modifiers of Systemic Lupus Erythematosus in *FcgammaRIIB(-/-)* Mice. *J. Exp. Med.* 195: 1167-1174.
- Bottomley, S. 1999. *Bioinformatics. Drug Discovery Today* 4: 482-484.
- Brandi, M. L., Gennari, L., Cerinic, M. M., Becherini, L., Falchetti, A., Masi, L., Gennari, C., and Reginster, J. Y. 2001. Genetic markers of osteoarticular disorders: facts and hopes. *Arthritis Res.* 3: 270-280.
- Brooks, P. C., Stromblad, S., Klemke, R., Visscher, D., Sarkar, F. H., and Cheresch, D. A. 1995. Antiintegrin alpha v beta 3 blocks human breast cancer growth and angiogenesis in human skin. *J. Clin. Invest.* 96: 1815-1822.
- Bumol, T. F. and Watanabe, A. M. 2001. Genetic information, genomic technologies, and the future of drug discovery. *J. Amer. Med. Assoc.* 285: 551-555.
- Cancedda, R., Castagnola, P., Cancedda, F. D., Dozin, B., and Quarto, R. 2000. Developmental control of chondrogenesis and osteogenesis. *Int. J. Dev. Biol.* 44: 707-714.
- Chen, Y., Dougherty, E.R., and Bittner, M. 1997. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Opt.* 2: 364-374.
- Childs, S., Chen, J. N., Garrity, D. M., and Fishman, M. C. 2002. Patterning of angiogenesis in the zebrafish embryo. *Development.* 129: 973-982.
- Chudin, E., Walker, R., Kosaka, A., Wu, S. X., Rabert, D., Chang, T. K., and Kreder, D. E. 2002. Assessment of the relationship between signal intensities and transcript concentration for Affymetrix GeneChip(R) arrays. *Genome Biol.* 3: RESEARCH0005.
- Cho, J. Y., Grant, T. D., Lunstrum, G. P., and Horton, W. A. 2001. Col2-GFP reporter mouse—A new tool to study skeletal development. *Am. J. Med. Genet.* 106: 251-253.
- Clark, A. G., Jordan, J. M., Vilim, V., Renner, J. B., Dragomir, A. D., Luta, G., and Kraus, V. B. 1999. Serum cartilage oligomeric matrix protein reflects osteoarthritis presence and severity: the Johnston County Osteoarthritis Project. *Arthritis Rheum.* 42: 2356-2364.

- Cronin, C. A., Gluba, W., and Scrable, H. 2001. The lac operator-repressor system is functional in the mouse. *Genes Dev.* 15: 1506-1517.
- Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., et al. 2002. A genomic regulatory network for development. *Science.* 295: 1669-1678.
- Denhardt, D. T., Noda, M., O'Regan, A. W., Pavlin, D., and Berman, J. S. 2001. Osteopontin as a means to cope with environmental insults: regulation of inflammation, tissue remodeling, and cell survival. *J. Clin. Invest.* 107: 1055-1061.
- Doerge, R. W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.* 3: 43-52.
- Doris, P. A. 2002. Hypertension genetics, single nucleotide polymorphisms, and the common disease: common variant hypothesis. *Hypertension.* 39: 323-331.
- Drews, J. 2000. Drug discovery: a historical perspective. *Science* 287: 1960-1964.
- Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* 95: 14863-14868.
- Ekholm, E., Hankenson, K. D., Uusitalo, H., Hiltunen, A., Gardner, H., Heino, J., and Penttinen, R. 2002. Diminished Callus Size and Cartilage Synthesis in alpha1beta1 Integrin-Deficient Mice during Bone Fracture Healing. *Am. J. Pathol.* 160: 1779-1785.
- Evans, C. H., Ghivizzani, S. C., Palmer, G. D., Gouze, J. N., Robbins, P. D., and Gouze, E. 2001. Gene therapy for rheumatoid arthritis. *Expert. Opin. Biol. Ther.* 1: 971-978.
- Fassler, R., Schnegelsberg, P. N., Dausman, J., Shinya, T., Muragaki, Y., McCarthy, M. T., Olsen, B. R., and Jaenisch, R. 1994. Mice lacking alpha 1 (IX) collagen develop noninflammatory degenerative joint disease. *Proc. Natl. Acad. Sci. U.S.A.* 91: 5070-5074.
- Forton, A. C., Petri, M. A., Goldman, D., and Sullivan, K. E. 2002. An osteopontin (SPP1) polymorphism is associated with systemic lupus erythematosus. *Hum. Mutat.* 19: 459-462.
- Frary, A., Nesbitt, T. C., Grandillo, S., Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K. B., and Tanksley, S. D. 2000. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science.* 289: 85-88.
- Freemont, A. J., Hampson, V., Tilman, R., Goupille, P., Taiwo, Y., and Hoyland, J. A. 1997. Gene expression of matrix metalloproteinases 1, 3, and 9 by chondrocytes in osteoarthritic human knee articular cartilage is zone and grade specific. *Ann. Rheum. Dis.* 56: 542-549.
- Ganu, V., M. R., Hu, S., Koehn, J., Klein, M., and Liebman, J. 2001. Extracellular matrix proteins COMP and fibronectin are substrates for Htra, a novel serine protease upregulated in osteoarthritic cartilage. 47th Annual meeting, Orthopaedic Research Society.
- Gerber, H. P., Vu, T. H., Ryan, A. M., Kowalski, J., Werb, Z., and Ferrara, N. 1999. VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation. *Nat. Med.* 5: 623-628.
- Ghivizzani, S. C., Kang, R., Georgescu, H. I., Lechman, E. R., Jaffurs, D., Engle, J. M., Watkins, S. C., Tindal, M. H., Suchanek, M. K., McKenzie, L. R., Evans, C. H., and Robbins, P. D. 1997. Constitutive intra-articular expression of human IL-1 beta following gene transfer to rabbit synovium produces all major pathologies of human rheumatoid arthritis. *J. Immunol.* 159: 3604-3612.
- Greely, H. T. 2001. Human genomics research. New challenges for research ethics. *Perspect. Biol. Med.* 44: 221-229.
- Hastie, T., Tibshirani, R., Eisen, M. B., Alizadeh, A., Levy, R., Staudt, L., Chan, W. C., Botstein, D., and Brown, P. 2000. 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. *Genome Biol.* 1: RESEARCH0003.
- Heinegard, D. and Oldberg, A. 1989. Structure and biology of cartilage and bone matrix noncollagenous macromolecules. *FASEB J.* 3: 2042-2051.
- Heinegard, D., Bayliss, M., and Lorenzo, P. 1998. Pathogenesis of Osteoarthritis. *Osteoarthritis Ed.* K.D. Brandt et al. Oxford University Press. p74-83.
- Hu, S. I., Carozza, M., Klein, M., Nantermet, P., Luk, D., and Crowl, R. M. 1998. Human Htra, an evolutionarily conserved serine protease identified as a differentially expressed gene product in osteoarthritic cartilage. *J. Biol. Chem.* 273: 34406-34412.
- Hurvitz, J. R., Suwairi, W. M., Van Hul, W., El Shanti, H., Superti-Furga, A., Roudier, J., Holderbaum, D., Pauli, R. M., Herd, J. K., Van Hul, E. V., Rezai-Delui, H., Legius, E., Le Merrer, M., Al Alami, J., Bahabri, S. A., and Warman, M. L. 1999. Mutations in the CCN gene family member WISP3 cause progressive pseudorheumatoid dysplasia. *Nat. Genet.* 23: 94-98.
- Ingvarsson, T., Stefansson, S. E., Gulcher, J. R., Jonsson, H. H., Jonsson, H., Frigge, M. L., Palsdottir, E., Olafsdottir, G., Jonsdottir, T., Walters, G. B., Lohmander, L. S., and Stefansson, K. 2001. A large Icelandic family with early osteoarthritis of the hip associated with a susceptibility locus on chromosome 16p. *Arthritis Rheum.* 44: 2548-2555.
- Ito, H., Akiyama, H., Iguchi, H., Iyama, K. K., Miyamoto, M., Ohsawa, K., and Nakamura, T. 2001. Molecular cloning and biological activity of a novel lysyl oxidase-related gene expressed in cartilage. *J. Biol. Chem.* 276: 24023-24029.
- Iyama, S., Okamoto, T., Sato, T., Yamauchi, N., Sato, Y., Sasaki, K., Takahashi, M., Tanaka, M., Adachi, T., Kogawa, K., Kato, J., Sakamaki, S., and Niitsu, Y. 2001. Treatment of murine collagen-induced arthritis by ex vivo extracellular superoxide dismutase gene transfer. *Arthritis Rheum.* 44: 2160-2167.
- Jawaheer, D., Seldin, M. F., Amos, C. I., Chen, W. V., et al., 2001. A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. *Am. J. Hum. Genet.* 68: 927-936.
- Jawaheer, D. and Gregersen, P. K. 2002. Rheumatoid arthritis. The genetic components. *Rheum. Dis. Clin. North Am.* 28:1-15.
- Jorgensen, C. and Gay, S. 1998. Gene therapy in osteoarticular diseases: where are we? *Immunol. Today.* 19: 387-391.
- Kimmel, C. B., Miller, C. T., Kruze, G., Ullmann, B., BreMiller, R. A., Larison, K. D., and Snyder, H. C. 1998.

- The shaping of pharyngeal cartilages during early development of the zebrafish. *Dev. Biol.* 203: 245-263.
- Kimmel, C. B., Miller, C. T., and Moens, C. B. 2001. Specification and morphogenesis of the zebrafish larval head skeleton. *Dev. Biol.* 233: 239-257.
- Kitano, H. 2002. Systems biology: a brief overview. *Science* 295: 1662-1664.
- Knight, J. 2001. When the chips are down. *Nature.* 410: 860-861.
- Koch, A. E. 1998. Review: angiogenesis: implications for rheumatoid arthritis. *Arthritis Rheum.* 41: 951-962.
- Kuivaniemi, H., Tromp, G., and Prockop, D. J. 1997. Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels. *Hum. Mutat.* 9: 300-315.
- Kumar, S., McDonnell, P. C., Lehr, R., Tierney, L., Tzimas, M. N., Griswold, D. E., Capper, E. A., Tal-Singer, R., Wells, G. I., Doyle, M. L., and Young, P. R. 2000. Identification and initial characterization of four novel members of the interleukin-1 family. *J. Biol. Chem.* 275: 10308-10314.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., et al., 2001. Initial sequencing and analysis of the human genome. *Nature.* 409: 860-921.
- Lin, H., Ho, A. S., Haley-Vicente, D., Zhang, J., Bernal-Fussell, J., Pace, A. M., Hansen, D., Schweighofer, K., Mize, N. K., and Ford, J. E. 2001. Cloning and characterization of il-1hy2, a novel interleukin-1 family member. *J. Biol. Chem.* 276: 20597-20602.
- Loeser, R. F. 2000. Chondrocyte integrin expression and function. *Biorheology* 37: 109-116.
- Luscombe, N. M., Greenbaum, D., and Gerstein, M. 2001. What is bioinformatics? A proposed definition and overview of the field. *Meth. Inf. Med.* 40: 346-358.
- Magill, G. 2002. The ethics weave in human genomics, embryonic stem cell research, and therapeutic cloning: promoting and protecting society's interests. *Albany. Law Rev.* 65: 701-728.
- Mahadevappa, M. and Warrington, J. A. 1999. A high-density probe array sample preparation method using 10- to 100-fold fewer cells. *Nat. Biotechnol.* 17: 1134-1136.
- Marcelino, J., Carpten, J. D., Suwairi, W. M., Gutierrez, O. M., et al., 1999. CACP, encoding a secreted proteoglycan, is mutated in camptodactyly-arthropathy-coxa vara-pericarditis syndrome. *Nat. Genet.* 23: 319-322.
- Martignetti, J. A., Aqeel, A. A., Sewairi, W. A., Boumah, C. E., Kambouris, M., Mayouf, S. A., Sheth, K. V., Eid, W. A., Dowling, O., Harris, J., Glucksman, M. J., Bahabri, S., Meyer, B. F., and Desnick, R. J. 2001. Mutation of the matrix metalloproteinase 2 gene (MMP2) causes a multicentric osteolysis and arthritis syndrome. *Nat. Genet.* 28: 261-265.
- Masinde, G. L., Li, X., Gu, W., Davidson, H., Mohan, S., and Baylink, D. J. 2001. Identification of wound healing/regeneration quantitative trait loci (QTL) at multiple time points that explain seventy percent of variance in (MRL/MpJ and SJL/J) mice F2 population. *Genome Res.* 11: 2027-2033.
- McCarty 1999. *Arthritis and Allied Conditions: A Textbook of Rheumatology*, 13th edn. Williams and Wilkins, Baltimore.
- Meulenbelt, I., Bijkerk, C., Miedema, H. S., Breedveld, F. C., Hofman, A., Valkenburg, H. A., Pols, H. A., Slagboom, P. E., and van Duijn, C. M. 1998. A genetic association study of the IGF-1 gene and radiological osteoarthritis in a population-based cohort study (the Rotterdam Study). *Ann. Rheum. Dis.* 57: 371-374.
- Mi, Z., Ghivizzani, S. C., Lechman, E. R., Jaffurs, D., Glorioso, J. C., Evans, C. H., and Robbins, P. D. 2000. Adenovirus-mediated gene transfer of insulin-like growth factor 1 stimulates proteoglycan synthesis in rabbit joints. *Arthritis Rheum.* 43: 2563-2570.
- Miller, C. T., Schilling, T. F., Lee, K., Parker, J., and Kimmel, C. B. 2000. sucker encodes a zebrafish Endothelin-1 required for ventral pharyngeal arch development. *Development.* 127: 3815-3828.
- Moos, V., Rudwaleit, M., Herzog, V., Hohlig, K., Sieper, J., and Muller, B. 2000. Association of genotypes affecting the expression of interleukin-1beta or interleukin-1 receptor antagonist with osteoarthritis. *Arthritis Rheum.* 43: 2417-2422.
- Moss, M. L., White, J. M., Lambert, M. H., and Andrews, R. C. 2001. TACE and other ADAM proteases as targets for drug discovery. *Drug Discovery Today.* 6: 417-426.
- Mott, R. and Flint, J. 2002. Simultaneous detection and fine mapping of quantitative trait Loci in mice using heterogeneous stocks. *Genetics* 160: 1609-1618.
- Mulero, J. J., Pace, A. M., Nelken, S. T., Loeb, D. B., Correa, T. R., Drmanac, R., and Ford, J. E. 1999. IL1HY1: A novel interleukin-1 receptor antagonist gene. *Biochem. Biophys. Res. Commun.* 263: 702-706.
- Nakata, K., Ono, K., Miyazaki, J., Olsen, B. R., Muragaki, Y., Adachi, E., Yamamura, K., and Kimura, T. 1993. Osteoarthritis associated with mild chondrodysplasia in transgenic mice expressing alpha 1(IX) collagen chains with a central deletion. *Proc. Natl. Acad. Sci. U.S.A* 90: 2870-2874.
- Newton, R. C., Solomon, K. A., Covington, M. B., Decicco, C. P., Haley, P. J., Friedman, S. M., and Vaddi, K. 2001. Biology of TACE inhibition. *Ann. Rheum. Dis.* 60 Suppl 3: 25-32.
- Noble, D. 2002. Modeling the heart—from genes to cells to the whole organ. *Science* 295: 1678-1682.
- Oates, A. C., Brownlie, A., Pratt, S. J., Irvine, D. V., Liao, E. C., Paw, B. H., Dorian, K. J., Johnson, S. L., Postlethwait, J. H., Zon, L. I., and Wilks, A. F. 1999. Gene duplication of zebrafish JAK2 homologs is accompanied by divergent embryonic expression patterns: only jak2a is expressed during erythropoiesis. *Blood* 94: 2622-2636.
- Patel, I. R., Attur, M. G., Patel, R. N., Stuchin, S. A., Abagyan, R. A., Abramson, S. B., and Amin, A. R. 1998. TNF- α convertase from human arthritis-affected cartilage: Isolation of cDNA by differential display, expression of the active enzyme, and regulation of TNF- α . *J. Immunol.* 160: 4570-4579.
- Perola, M., Ohman, M., Hiekkalinna, T., Leppavuori, J., Pajukanta, P., Wessman, M., Koskenvuo, M., Palotie, A., Lange, K., Kaprio, J., and Peltonen, L. 2001. Quantitative-trait-locus analysis of body-mass index and of stature, by combined analysis of genome scans of five Finnish study groups. *Am. J. Hum. Genet.* 69: 117-123.
- Petit, B., Freyria, A. M., van der Rest, M., and Herbage, D.

1992. Cartilage collagens. In *Biological Regulation of the Chondrocytes* Ed: M. Adolphe. pp. 33-84.
- Piedra, M. E., Delgado, M. D., Ros, M. A., and Leon, J. 2002. c-Myc overexpression increases cell size and impairs cartilage differentiation during chick limb development. *Cell Growth Differ.* 13: 185-193.
- Pullig, O., Weseloh, G., Gauer, S., and Swoboda, B. 2000. Osteopontin is expressed by adult human osteoarthritic chondrocytes: protein and mRNA analysis of normal and osteoarthritic cartilage. *Matrix Biol.* 19: 245-255.
- Quackenbush, J. 2001. Computational analysis of microarray data. *Nat. Rev. Genet.* 2: 418-427.
- Rivera, M. C., Jain, R., Moore, J. E., and Lake, J. A. 1998. Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. U.S.A.* 95: 6239-6244.
- Rozzo, S. J., Allard, J. D., Choubey, D., Vyse, T. J., Izui, S., Peltz, G., and Kotzin, B. L. 2001. Evidence for an interferon-inducible gene, Ifi202, in the susceptibility to systemic lupus. *Immunity.* 15: 435-443.
- Sachdev, S. W., Dietz, U. H., Oshima, Y., Lang, M. R., Knapik, E. W., Hiraki, Y., and Shukunami, C. 2001. Sequence analysis of zebrafish chondromodulin-1 and expression profile in the notochord and chondrogenic regions during cartilage morphogenesis. *Mech. Dev.* 105: 157-162.
- Saito, S., Yamaji, N., Yasunaga, K., Saito, T., Matsumoto, S., Katoh, M., Kobayashi, S., and Masuho, Y. 1999. The fibronectin extra domain A activates matrix metalloproteinase gene expression by an interleukin-1-dependent mechanism. *J. Biol. Chem.* 274: 30756-30763.
- Schmidt, S., Barcellos, L. F., DeSombre, K., Rimmler, J. B., Lincoln, R. R., Bucher, P., Saunders, A. M., Lai, E., Martin, E. R., Vance, J. M., Oksenberg, J. R., Hauser, S. L., Pericak-Vance, M. A., and Haines, J. L. 2002. Association of polymorphisms in the apolipoprotein E region with susceptibility to and progression of multiple sclerosis. *Am. J. Hum. Genet.* 70: 708-717.
- Schnabel, M., Marlovits, S., Eckhoff, G., Fichtel, I., Gotzen, L., Vecsei, V., and Schlegel, J. 2002. Dedifferentiation-associated changes in morphology and gene expression in primary human articular chondrocytes in cell culture. *Osteoarthritis and Cartilage* 10: 62-70.
- Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* 95: 5857-5864.
- Scola, M. P., Imagawa, T., Boivin, G. P., Giannini, E. H., Glass, D. N., Hirsch, R., and Grom, A. A. 2001. Expression of angiogenic factors in juvenile rheumatoid arthritis: correlation with revascularization of human synovium engrafted into SCID mice. *Arthritis Rheum.* 44: 794-801.
- Scott, B. B., Zaratini, P. F., Colombo, A., Hansbury, M. J., Winkler, J. D., and Jackson, J. R. 2002. Constitutive expression of angiopoietin-1 and -2 and modulation of their expression by inflammatory cytokines in rheumatoid arthritis synovial fibroblasts. *J. Rheumatol.* 29: 230-239.
- Scully, S. P., Lee, J. W., Ghert, P. M. A., and Qi, W. 2001. The role of the extracellular matrix in articular chondrocyte regulation. *Clin. Orthop.* S72-S89.
- Silman, A. J. 2002. Commentary: Do genes or environment influence development of rheumatoid arthritis? *Brit. Med. J.* 324:264.
- Smith, P., Shuler, F. D., Georgescu, H. I., Ghivizzani, S. C., Johnstone, B., Niyibizi, C., Robbins, P. D., and Evans, C. H. 2000. Genetic enhancement of matrix synthesis by articular chondrocytes: comparison of different growth factor genes in the presence and absence of interleukin-1. *Arthritis Rheum.* 43: 1156-1164.
- Sonnhammer, E. L., Eddy, S. R., and Durbin, R. 1997. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins.* 28: 405-420.
- Stokes, D. G., Liu, G., Coimbra, I. B., Piera-Velazquez, S., Crowl, R. M., and Jimenez, S. A. 2002. Assessment of the gene expression profile of differentiated and dedifferentiated human fetal chondrocytes by microarray analysis. *Arthritis Rheum.* 46: 404-419.
- Storgard, C. M., Stupack, D. G., Jonczyk, A., Goodman, S. L., Fox, R. I., and Cheres, D. A. 1999. Decreased angiogenesis and arthritic disease in rabbits treated with an alphavbeta3 antagonist. *J. Clin. Invest* 103: 47-54.
- Strohman, R. 2002. Maneuvering in the complex path from genotype to phenotype. *Science* 296: 701-703.
- Stuhlmuller, B., Ungethum, U., Scholze, S., Martinez, L., Backhaus, M., Kraetsch, H. G., Kinne, R. W., and Burmester, G. R. 2000. Identification of known and novel genes in activated monocytes from patients with rheumatoid arthritis. *Arthritis Rheum.* 43: 775-790.
- Subramanian, G., Adams, M. D., Venter, J. C., and Broder, S. 2001. Implications of the human genome for understanding human biology and medicine. *J. Amer. Med. Assoc.* 286: 2296-2307.
- Svensen, A. J., Holm, N. V., Kyvik, K., Petersen, P. H., and Junker, P. 2002. Relative importance of genetic effects in rheumatoid arthritis: historical cohort study of Danish nationwide twin population. *Brit. Med. J.* 324: 264-266.
- Symul, D. J., Frazer, K. A., Ueda, Y., Deneffle, P., Stevens, M. E., Wang, Z. E., Locksley, R., and Rubin, E. M. 1999. Functional screening of an asthma QTL in YAC transgenic mice. *Nat. Genet.* 23: 241-244.
- Taniguchi, K., Kohsaka, H., Inoue, N., Terada, Y., Ito, H., Hirokawa, K., and Miyasaka, N. 1999. Induction of the p16INK4a senescence gene as a new therapeutic strategy for the treatment of rheumatoid arthritis. *Nat. Med.* 5: 760-767.
- Teraoka, H., Dong, W., Ogawa, S., Tsukiyama, S., Okuhara, Y., Niiyama, M., Ueno, N., Peterson, R. E., and Hiraga, T. 2002. 2,3,7,8-Tetrachlorodibenzo-p-dioxin toxicity in the zebrafish embryo: altered regional blood flow and impaired lower jaw development. *Toxicol. Sci.* 65: 192-199.
- Tetlow, L. C., Adlam, D. J., and Woolley, D. E. 2001. Matrix metalloproteinase and proinflammatory cytokine production by chondrocytes of human osteoarthritic cartilage: associations with degenerative changes. *Arthritis Rheum.* 44: 585-594.
- Van den Berg, W. B. 1995. Growth factors in experimental osteoarthritis: transforming growth factor beta pathogenic? *J. Rheumatol. Suppl* 43: 143-145.
- van Meurs, J., van Lent, P., Holthuysen, A., Lambrou, D., Bayne, E., Singer, I., and van den, B. W. 1999. Active

- matrix metalloproteinases are present in cartilage during immune complex-mediated arthritis: a pivotal role for stromelysin-1 in cartilage destruction. *J. Immunol.* 163: 5633-5639.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., *et al.*, 2001. The sequence of the human genome. *Science* 291: 1304-1351.
- Wang, F. L., Connor, J. R., Dodds, R. A., James, I. E., Kumar, S., Zou, C., Lark, M. W., Gowen, M., and Nuttall, M. E. 2000. Differential expression of *egr-1* in osteoarthritic compared to normal adult human articular cartilage. *Osteoarthritis and Cartilage*. 8: 161-169.
- Warrington, J. A., Nair, A., Mahadevappa, M., and Tsyganskaya, M. 2000. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol. Genomics* 2: 143-147.
- Wolfsberg, T. G., Bazan, J. F., Blobel, C. P., Myles, D. G., Primakoff, P., and White, J. M. 1993. The precursor region of a protein active in sperm-egg fusion contains a metalloprotease and a disintegrin domain: structural, functional, and evolutionary implications. *Proc. Natl. Acad. Sci. U.S.A.* 90: 10783-10787.
- Woods, J. M., Katschke, K. J., Volin, M. V., Ruth, J. H., Woodruff, D. C., Amin, M. A., Connors, M. A., Kurata, H., Arai, K., Haines, G. K., Kumar, P., and Koch, A. E. 2001. IL-4 adenoviral gene therapy reduces inflammation, proinflammatory cytokines, vascularization, and bony destruction in rat adjuvant-induced arthritis. *J. Immunol.* 166: 1214-1222.
- Wright, G. D., Hughes, A. E., Regan, M., and Doherty, M. 1996. Association of two loci on chromosome 2q with nodal osteoarthritis. *Ann. Rheum. Dis.* 55: 317-319.
- Wu, T. D. 2001. Analysing gene expression data from DNA microarrays to identify candidate genes. *J. Pathol.* 195: 53-65.
- Yamada, Y., Okuizumi, H., Miyauchi, A., Takagi, Y., Ikeda, K., and Harada, A. 2000. Association of transforming growth factor β 1 genotype with spinal osteophytosis in Japanese women. *Arthritis Rheum.* 43: 452-460.
- Yan, Y. L., Jowett, T., and Postlethwait, J. H. 1998. Ectopic expression of *hoxb2* after retinoic acid treatment or mRNA injection: disruption of hindbrain and craniofacial morphogenesis in zebrafish embryos. *Dev. Dyn.* 213: 370-385.
- Yaspo, M. L. 2001. Taking a functional genomics approach in molecular medicine. *Trends Mol. Med.* 7: 494-501.
- Zhou, Z., Apte, S. S., Soininen, R., Cao, R., Baaklini, G. Y., Rauser, R. W., Wang, J., Cao, Y., and Tryggvason, K. 2000. Impaired endochondral ossification and angiogenesis in mice deficient in membrane-type matrix metalloproteinase I. *Proc. Natl. Acad. Sci. U.S.A.* 97: 4052-4057.