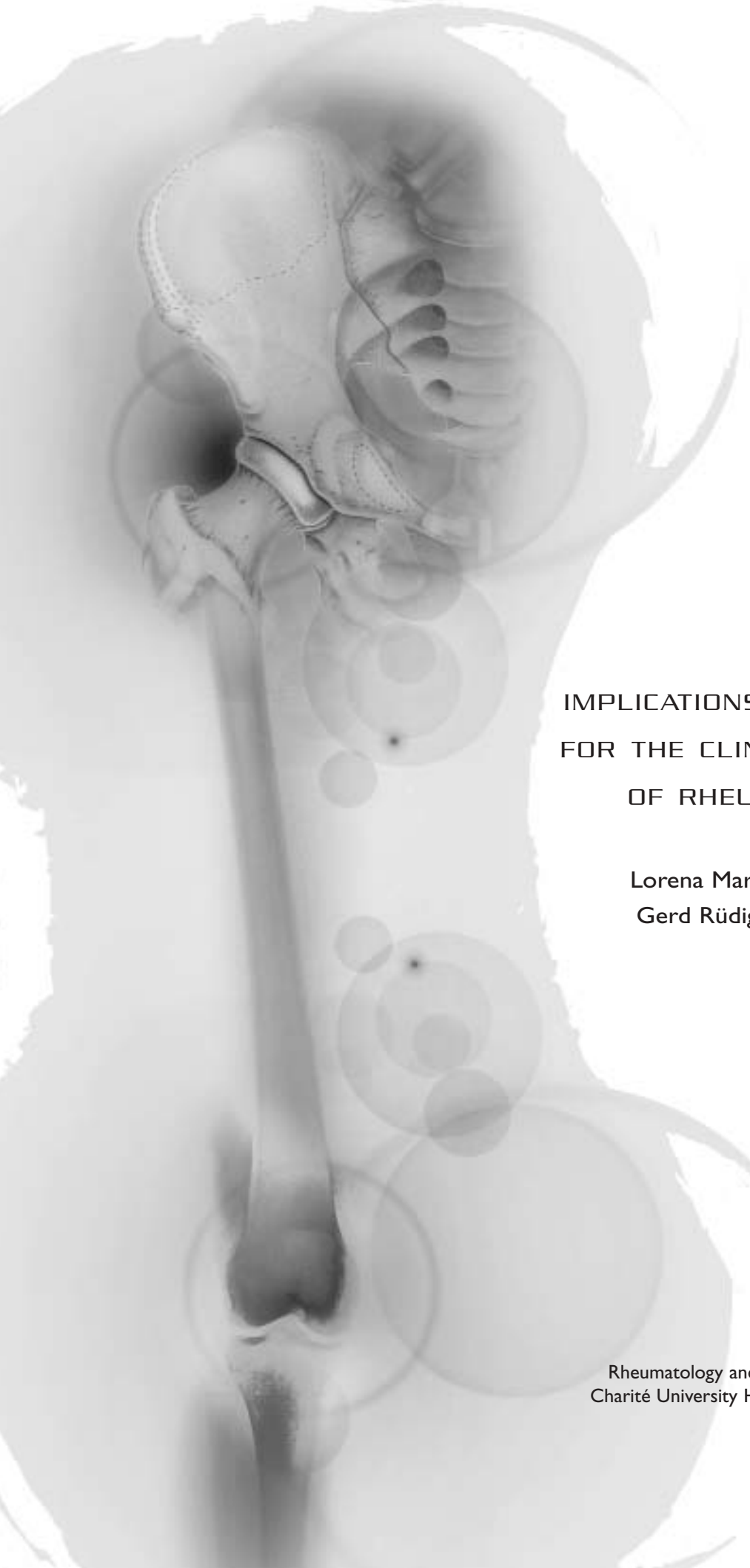

IMPLICATIONS OF GENOMICS FOR THE CLINICAL PRACTICE OF RHEUMATOLOGY

Lorena Martinez Gamboa,
Gerd Rüdiger Burmester

Rheumatology and Clinical Immunology,
Charité University Hospital, Berlin, Germany

ABSTRACT

In the era of the "genomic revolution", three major disciplines are widely recognized: the original genomics, and the directly related transcriptomics and proteomics. A fundamental aspect for development of these and other disciplines is the expansion from studying single genes or proteins to studying all at once in a systematic way, in the so called genome-wide approaches. As tool for discovery, these approaches are rapidly gaining influence in all areas of medical sciences, including Rheumatology.

This review gives an overview about genomics and its relationship with other related "-omics" disciplines, and discuss some advantages and limitations when working with genome-wide approaches. Furthemore, several examples are presented on applications of these technologies to the study of some rheumatic disorders. Finally, perspectives and potential clinical application for gene or protein candidates resulting from such approaches are discussed.

In order to accelerate comprehensive interpretation of obtained data, and also to give all interested groups worldwide the opportunity of participating in this fascinating discovery process, an appeal is done for encouraging cooperation and exchange of information between groups and, ideally, for allowing accessibility to all obtained information from genome-wide approaches at all levels (DNA, RNA, proteins and beyond) in public databases.

Keywords: Genomics; Proteomics; Rheumatic diseases.

RESUMO

Na era da «revolução genómica», três disciplinas principais são largamente reconhecidas: a genómica original e as directamente relacionadas transcriptómica e proteómica. Um aspecto fundamental para o desenvolvimento destas e de outras disciplinas é expandir o estudo de genes únicos ou de proteínas, passando a estudar os mesmos unitária e sistematicamente, através de procedimentos que envolvem o genoma como um todo. Como ferramenta que favorece a descoberta, estes procedimentos estão ganhando rapidamente influência em todas as áreas das Ciências Médicas, inclusive na Reumatologia.

Esta revisão fornece um resumo geral sobre a genómica e sua relação com outras disciplinas «-ómicas» afins, e discute algumas vantagens e limitações do trabalho com procedimentos que envolvem todo o genoma. Além disso, são apresentados vários exemplos de aplicações destas áreas no estudo de algumas doenças reumáticas. Por fim, são discutidas perspectivas e possíveis aplicacões clínicas para os genes e proteínas candidatos, resultantes de tais procedimentos.

De forma a obter uma interpretacão eficaz e rápida dos dados obtidos e também dar a todos os grupos interessados, em todo o mundo, a oportunidade de participar deste processo fascinante de descoberta, é feito um apelo em prol da cooperação e da troca de informação entre grupos e, o que seria ideal, da permissão de acesso a toda a informação obtida através de procedimentos envolvendo o genoma como um todo, em todos os níveis (DNA, RNA, proteínas e outros), em bancos de dados públicos.

Palavras-Chave: Genómica; Proteómica; Doenças reumáticas.

IMPLICATIONS OF GENOMICS FOR THE CLINICAL PRACTICE OF RHEUMATOLOGY

Lorena Martinez Gamboa*,
Gerd Rüdiger Burmester**

Introduction

Rheumatic diseases are the most prevalent chronic inflammatory disorders and an important cause of disability, encompassing more than 100 different entities and conditions, which implicate an enormous social, psychological and economical impact[1]. Despite many efforts done through the last decades in order to learn more about these diseases, it is still relatively little known about their aethiopathogenesis. However, it is recognized that for initiation of rheumatic autoimmune processes, a conjunction of external factors must encounter genetic susceptible hosts. Subsequent activation of the immune system leads to inflammation, abnormal regulation of the autoimmune response, perpetuation of the inflammatory process, and finally, to destruction of involved tissues, resulting in irreversible organ damage and death.

In the era of genomics, genome-wide approaches are rapidly gaining influence in all areas of medical sciences, including Rheumatology. This sort of analysis on differential gene expression and dysfunction applied to rheumatic diseases could help find the lost link between the triggering event(s) and the initiation and development of such disorders, and consequently, help understand the molecular mechanisms underlying these pathologic processes.

The objective of this review is to give an overview about genomics and its relationships with other related "-omics" disciplines, and their potential implications for the clinical practice of Rheumatology.

*Biologist, Scientist. Charite University Hospital – Campus Mitte. Medical Clinic III for Rheumatology, Clinical Immunology and Osteology. Laboratory # 329 at the German Center for Investigations on Rheumatic Diseases (Deutsches Rheuma-Forschungszentrum Berlin, DRFZ)
**Medical Doctor, Director of the Clinic. Charite University Hospital – Campus Mitte. Medical Clinic III for Rheumatology, Clinical Immunology and Osteology. Medical Faculty – Humboldt University of Berlin.

About Omics And Omes: Genomics and Related Disciplines

"Omics" is a neologism used as suffix attached to the name of biological subjects, e.g. genomics, transcriptomics, proteomics, etc. It is used for describing very large-scale data collection and analysis, meaning the study of a whole "object". The "objects" in this case are the "omes", from the Greek for "all", "every" or "complete": genome, transcriptome, proteome, etc. Actually, the original use of the suffix "ome" was in the word "genome", which refers to the complete genetic information of an organism. Because of the success of large-scale biology projects, the suffix "ome" has been extended to other contexts. It should be noticed that while the new term "genomics" deals with the global properties of genomes, the old term "genetics" generally refers to the analysis of single genes or groups of genes[2].

In the years preceding elucidation of the complete DNA sequence from human and other organisms, the concept of genomics was used to describe the new emerging discipline dealing with mapping and sequencing of genomes[3]. The American College of Rheumatology, recognizing the importance of this new discipline for Rheumatology, organized the *2000 Basic Research Conference: Genetics and Genomics in Rheumatic Diseases*, held as preface to the ACR 2000 Annual Meeting in Philadelphia[4]. Here, a first prospective analysis was made about possible implications of these new technologies on patient care.

In June 2000, the Human Genome Project (HGP) and Celera officially announced completion of a working draft of the human genome sequence[5;6]. Finally, in February 2001, the publication of the genome sequence was announced: the public sequence from the HGP in Nature and the Celera sequence in Science[7;8].

After completion of sequencing of the human genome, in the so called post-genomic era, the term "genomics" has evolved, and can now be seen in a

broader sense as a discovery process aimed at analysis and comprehension of the genetic information, which can be divided in two phases: structural and functional genomics. In this context, structural genomics comprises the initial phase of genome analysis, which is aimed at defining the structure and organization of the complete DNA sequence as the ultimate physical map of an organism[9]. For human genomics, this goal was reached after finishing the genome sequence. In the second phase, making use of the information provided by structural genomics, functional genomics refers to development and application of experimental approaches to assess gene expression. However, biological functions and phenotypic manifestations of genes can not be elucidated based solely on genetic analysis. Due to the broad range and the quantity of emerging new information covering several topics of scientific knowledge, this information cannot longer be managed by a single discipline. For this reason, several related scientific areas, the majority of them added with the magical suffix "-omics", emerged in the last years, each one of them devoting to a specific aspect of functional genomics. Some of them are explained next. An overview about the interactions of these related disciplines in the context of genomics is schematically presented in Figure 1.

Genomic polymorphism analysis

A part of functional genomics deals directly with search for and analysis of functional polymorphisms and mutations at the DNA level. Besides classical sequencing of regions of interest, SNP (for single nucleotide polymorphisms) approaches search for variations in single bases of the DNA. SNPs comprise approximately 80% of all known polymorphisms in the human genome and their occurrence is estimated to be 1 in 1.000 base pairs[10]. Used for large-scale genome-wide association studies, the hope is to find SNPs within or in the vicinity of a gene directly involved in a disease and in this way, to use them as genetic markers[11]. Neverthless, due to the magnitude of the human genome, analysis of single nucleotide polymorphisms can not be done for all the genome at the same time. Rather, reported analyses on SNPs usually refer to one or several polymorphisms present in one or more genes involved in a specific disease. Therefore, although SNP analysis can be considered as a part of functional genomics, speaking in a strict sense it does not represent a genomic approach. Nevertheless, SNP analysis is considered here because its importance for the contribution to the understanding of genetic mechanisms of diseases at the DNA level.

Transcriptomics

Gene expression is analyzed at the RNA level by measurement of transcripts within the area of transcriptomics. The methodology of choice nowadays is the micro-array technology. This is based on the hybridization of transcripts in a clinical sample to labeled cDNAs, PCR-products or synthetic DNA oligonucleotides, which are bound to a glass slide in a co-ordinate system, so called "array"[12]. In this way, a snapshot of the gene activity at a given time point can be obtained.

Proteomics

Furthermore, translated RNA messages are assessed at the protein level in the area of proteomics. The development of this global analysis of all proteins present in a given sample was achieved principally through the conjunction of two methods: two-dimensional gel electrophoresis (2-DE) and mass spectrometry (MS)[13;14]. Dynamic changes in a sample may have a high level of complexity and could be caused by altered synthesis of proteins, protein degradation, or post-translational modifications. Thus, they cannot be evaluated at the genomic or transcriptomic levels, but rather can be assessed by proteomic approaches[15].

Beyond these three relative well-known and well-established areas, a couple of related, new "-omics" disciplines have emerged in the last years and are still emerging, which deal with the final goal of comprehension of gene function from different points of view. Two of these other "omics" disciplines are worthy of mention here because of their potential implications for rheumatic diseases, and should be explained shortly: immunomics and metabolomics.

Immunomics

The first one, immunomics, refers to the study of immune-related transcripts, in order to unravel the molecular functions, regulation and diversity of immunoglobulins, receptors, cytokines and other immune system components[16]. For example, auto-antibody profiles have been determined with antigen-based microarrays[17], and cyto- and chemokine profiles with antibody-based arrays[18].

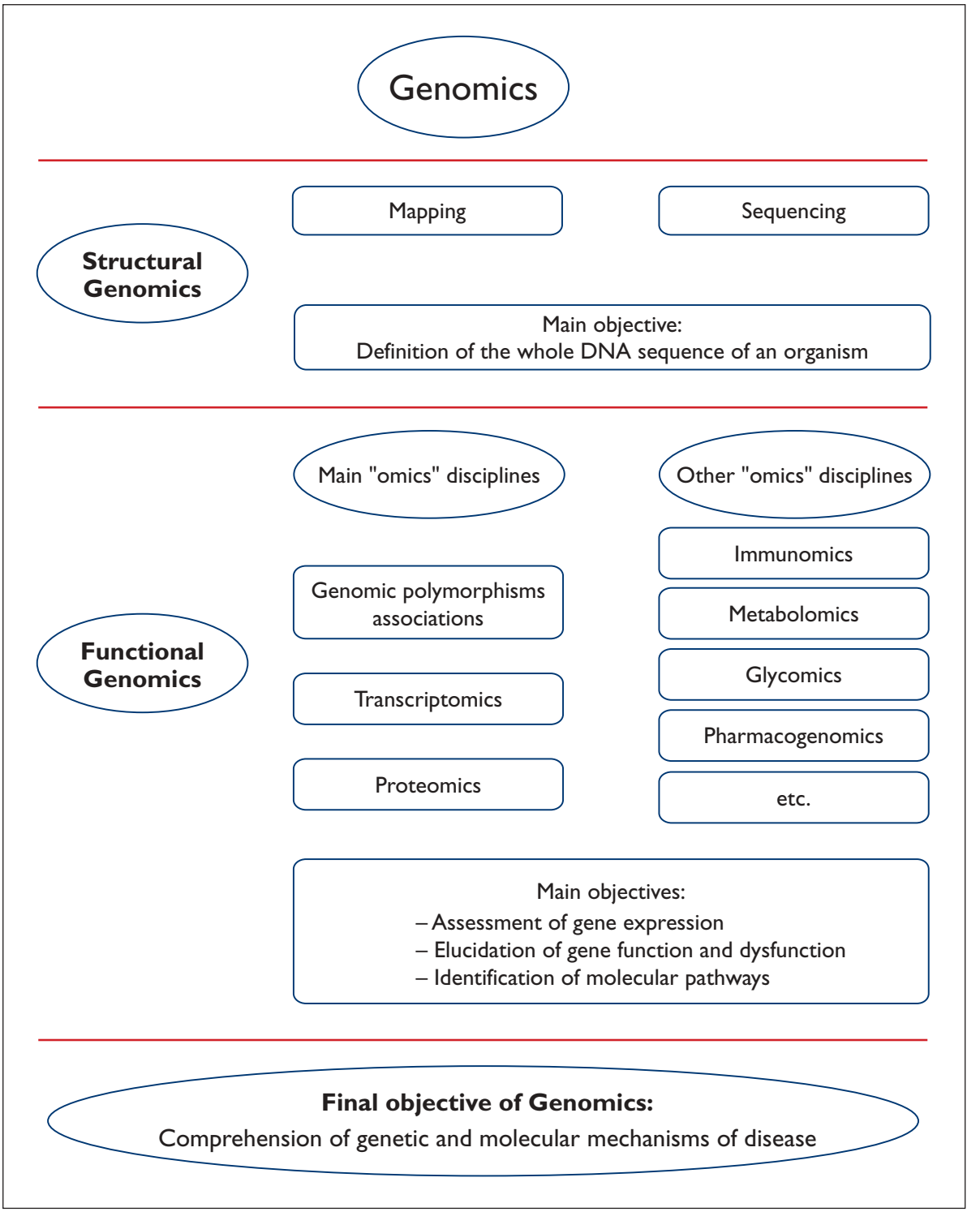

Figure 1. Overview of genomics and related disciplines as tools for the comprehension of genetic mechanisms of diseases.

These methodologies represent an accurate low-cost and low-sample option for simultaneous determination of several immunological targets, that otherwise had to be determined one by one in single tests.

Metabolomics

The second one, metabolomics, is the area concerned with analysis of global profiles of low molecular weight (MW) metabolites. The low MW feature of the metabolome constitutes the counterpart of the proteome, which deals with high MW proteins. The profiles are generated using methods such as mass spectrometry and nuclear magnetic resonance[19;20]. Experts in this area opine that, since the metabolome reflects an answer of an organism to genetic alterations, environmental influences or pathological processes, metabolite profiles of body fluids or tissues can be important indicators of physiological or pathological states[21;22].

One fundamental aspect in the development of these and other disciplines is the expansion from studying single genes or proteins to studying all at once in a systematic way, in the so called genome-wide approaches. Analysis and management of this overwhelming flood of data and its assimilation in the context of these related disciplines would not be possible without strong support from bioinformatic sciences. Besides evaluation and assistance in the interpretation of data, computational tasks can also include: establishment and management of databases, mathematical and structure modeling, simulation of laboratory experiments and prediction of experimental outcomes and statistical support[23].

Gradually, the conjunction of these enormous efforts should lead to assessment of gene expression, elucidation of gene function and identification of gene dysfunction, with the final purpose of a definite comprehension of the genetic and molecular mechanisms of disease.

Possibilities And Limitations of Genomic Approaches Applied to Rheumatic Diseases

The first argument supporting the realization of genome-wide functional analysis related to rheumatic diseases is the still relative small knowledge on genetic and molecular mechanisms underlying these disorders, in spite of their huge social, psychological and economical impact. Therefore, each attempt to find out more about the dysfunctions dominating these diseases and how to treat them in benefit of the patients should be encouraged.

Study populations

The relatively high frequency of these diseases constitutes an important advantage for the achievement of functional studies. Cohorts with appropriate numbers of patients and with defined disease parameters, adequate for powerful statistical analysis, can be established in rather short time. In this way, studies about actual disease activation phases as well as comparisons before and after a defined treatment can easily be performed. Furthermore, familiar studies for analysis of inheritance related aspects can also be accomplished.

Clinical samples

Another great advantage when investigating rheumatic diseases may be the feasibility in the acquisition of adequate clinical samples. The ideal set up in functional studies is to analyze single cell populations. Pure peripheral blood cell populations, e.g. CD4+ or CD8+ T-lymphocytes, CD14+ monocytes, CD19+ B-cells, etc., can be readily obtained from peripheral whole blood by well-established methods. Additionally, in some cases, access to affected tissues is easily practicable, as is the case of joint tissues in arthritis. Nevertheless, separation of a specific cell type from an affected tissue or organ (e.g. lung, kidney, infiltrated joints, etc.) could be sometimes problematic. In cases where mixed cell populations are analyzed, statements are done for tissue inflammation status. In any case, it is recommended to thoroughly plan processing of clinical samples in order to obtain reliable information on function and/or activity of the one cell type or tissue of interest.

Infrastructure

Due to the comparative high processing costs and the special infrastructure needed for application of these high-throughput technologies, the recently developed but differently implemented platforms are not utilized routinely, making difficult to compare the experimental approaches and the results obtained. For their use in the routine laboratory analysis, these methodologies need to be standardized and ranges of reference need to be defined. Until this is accomplished, it is important to emphasize that comparisons between different expe-

riments are only allowed when performed with the same system and under several intra- and inter-experimental internal controls.

Bioinformatics

As mentioned before, the management of this enormous amount of information would not be possible without assistance from computational sciences. Informatic professionals who are in charge of analyzing the results from genome-wide approaches have the advantage of impartiality, since their interpretation of the data is solely based on the result of a formula or an algorithm, so that in their eyes all genes are "equal". However, the decision on a "positive" or "negative" regulated gene cannot be taken only based on signal intensity thresholds, an algorithm, or other mathematical parameters. Rather, for doing this classification, known biological principles and molecular processes must also be taken into account, which, in the best of the cases, should be compatible with the clinical observations. Therefore, it is of pivotal importance to promote formation of computational professionals with a deeper knowledge of biological processes on the one hand, and to encourage the development of "scientists-friendly" software packages on the other hand, in order to allow a tightly and fruitfully cooperation between biologists and bioinformatic professionals.

Validation

Extremely important is the fact that results from statistical analysis of discovery approaches at any level (DNA, RNA, protein, antibodies, etc.) must be validated, that means, be reproduced. Nevertheless, due to the time and the extensive experiments required for validation analysis, only a limited set of selected candidate genes or biomarkers can be further evaluated. As an example of this can be mentioned the work by Liao and collegues, which needed 1 month per sample for the validation experiments using the same methodology as the discovery approaches[15;24]. Consequently, besides the limited number of candidates which can be validated, their selection for extensive analysis depends ultimately on the research focus of the investigating group.

Interpretation

Since these methodologies are not applied routinely, it is still a sort of "privilege" of only several groups worldwide to perform these kinds of genome-wide approaches and to select the information to be published. In this sense, only a small percentage of the information obtained on gene or protein expression approaches is really used for publication and for additional work, depending, as mentioned before, on the special interests of the investigating group. A great part of the data will probably remain in a drawer, since no single group can pretend to further investigate all differentially expressed genes in a study cohort. Unfortunately, that means first, that the time as also the personal and economical resources invested in such approaches remain sub-utilized, and second, that the information reaching the public community is limited to the scope of the publishing group.

What do We Have Until Now?

The nucleotide sequence of many human genes has been already defined, either through prediction from the DNA sequence or through isolation and identification from a clinical sample. Similarly, the corresponding proteins have been either elucidated from the DNA sequence or have been actually isolated. Nevertheless, elucidation of a biological function and assignment to a pathway has been achieved only for a small percentage of them[25]. Contrary to the many expectations on the beginnings of the genomic revolution, genomic approaches have brought up to now only first suggestion of candidates which could be specifically involved in a pathological process. In almost all cases, further investigations are necessary for verification of a real pathological contribution to the disease.

Various reports on different discovery approaches applied to some rheumatic diseases are presented next. A more extensive relation on the current status of gene expression profiling in rheumatic diseases has been discussed for example by Häupl *et al.*[26].

From Analysis of Single Nucleotide Polymorphisms

Over the last years, there have been many reports on SNPs analysis for all thinkable genes and diseases. However, initial expectations of finding powerful disease-related polymorphisms have diminished with appearance of contradictory results. It seems to be that the most common SNPs may not be the most informative[27]. According to the group of Chakravarti[28], most SNPs may not have a direct influence on their protein products, because they

fall in the non-coding area of the genome. Furthermore, it seems to be that a significant percentage of relevant SNP variants occur principally in certain populations, such as Asians or African-Americans. In order to discover relevant SNPs, sample populations should be large and ethnically diverse. Additionally, to detect disease-relevant SNPs, patients' cohorts should be as large as possible.

These difficulties in the analysis of SNPs are reflected in the inconsistent reports for several genes in several populations. A good example for this is the analysis of the TNF promoter polymorphisms, which have been analyzed for example in patients with RA[29-33], SLE [34] and psoriasis arthritis[35]. The results from these analysis are miscellaneous, showing in most of the cases little consensus on whether any TNF polymorphisms are in fact functionally associated to the diseases[36].

Other examples of genes with SNPs apparently associated to rheumatic diseases, which may be worthy of further investigation, are: for RA, IL-1ß[37]; for SLE, osteopontin[38], IL-1ß [37] and the B-cell marker CD19[39]; and for SSc: IL-1A[40], SPARC[41] and fibrillin 1[42;43].

From transcriptomic approaches

Heller *et al.*[44] made a first report on the use of a customized cDNA microarray for the analysis of gene expression in patients with RA. For their array, they selected a set of human genes of probable significance in inflammation, and tested it with synovial tissue from RA patients and, for comparison, with intestinal tissue samples from patients with Crohn's disease, cultured chondrocyte cell lines, primary chondrocytes, and synoviocytes. Among others, comparisons allowed them to describe differential expression of tissue inhibitor of metalloproteinase 1, ferritin light chain, and manganese superoxide dismutase in RA when compared to Crohn's disease. Based on their results, they demonstrate the suitability of cDNA micro-array technology for profiling of complex diseases.

An interesting study was reported last year by Bovin and colleagues[45]. They compared the gene expression in peripheral blood mononuclear cells from rheumatoid factor (RF) – positive with negative RA patients, and also with healthy donors (HD), in oligonucleotide-based micro-arrays. After analysis and validation, they found no significant differences in PBMC expression profiles from seropositive and seronegative patients, suggesting that RF-positive and RF-negative RA are not fundamentally different diseases. Furthermore, when comparing RA patients with HD, they found following genes having a significantly higher expression in RA patients than in HD: CD14 antigen, defensin alpha-1 and alpha-3 (DEFA), fatty-acid--Coenzyme A ligase, long-chain 2 (FACL), ribonuclease 2 (RNASE2), and the calcium-binding proteins S100A8 and S100A12 (calgranulin A and C). The last two have been also reported to be up-regulated at the protein level, as will be mentioned later[24]. Additionally, the expression of HLA-DQ beta 1 was significantly reduced in RA patients compared to HD. Genes for MHC class II molecules, DP alpha 1 and DR alpha, have also been found in another study to be down-regulated in early RA[46].

The newest work published up to now is that reported by Olsen *et al.*[46]. Here, they compared gene expression profiles of peripheral blood mononuclear cells from patients with recent onset RA (ERA) and with longstanding RA (LRA), and compared them also with patients with SLE, asthma, allergic diseases, and healthy donors, on commercial micro-arrays. Remarkable, they found 44 genes that were down regulated by more than threefold in ERA when compared to LRA. These genes were grouped into functional categories, from which the two major groups comprised genes involved in immune and inflammatory responses and cell proliferation. Furthermore, they refer about nine genes up regulated in ERA, including three involved in activities of the immune system (TGFß receptor II, CSF3 receptor, cleavage stimulating factor) and two involved in glucocorticoids activity (cytochrome P450 IIIA, 11-ß hydroxysteroid dehydrogenase 2). In conclusion, the authors suggest that early RA presents a characteristic gene expression profile, and emphasized the importance of early and accurate diagnosis of RA for avoiding long term disability, thus pointing out the applicability of their results.

From Proteomic Approaches

Using proteomic approaches, Dotzlaw *et al.*[47] were able to establish a protein expression profile in peripheral blood mononuclear cells of RA patients, different from that for healthy donors. They identified a total of 29 differentially over- or under-expressed proteins in RA patients when compared to healthy controls. Among the up-regulated proteins were HSPA5, a demonstrated autoantigen in RA[48], and fibrinogen gamma, which could have a potential role in inflammatory arthritis through acti-

vation of chemokine production in synovial fibroblasts[49]. On the other side, the heat shock protein Hsp60, which has been shown to contribute to arthritis suppression by stimulation of regulatory suppressive T cell activity[50], was found to be down-regulated[51]. However, the question remains open whether these proteins are directly involved in the pathogenesis of RA, or are gene products with altered expression as consequence of the disease.

In another proteomic study, Liao *et al.*[24] described the search for differentially expressed protein biomarkers when analyzing synovial fluid (SF) in patients with erosive and non-erosive RA. From a total of 418 identified proteins, they selected 30 that were more abundant in SF from patients with erosive RA and 3 that were more abundant in non-erosive RA SF. Selected candidates included serum proteins, metabolic enzymes, calcium-binding S100 proteins, matrix-degrading proteinases, cellular signaling proteins and several not-yet characterized proteins. For validation, they selected CRP as control, because of its feasibility for measurement, and six members of the S100 family of calcium-binding proteins (calgranulin A -S100A8-, B -S100A9- and C -S100A12-, calgizzarin -S100A11-, -metastasin -S100A4- and S100P protein). Interestingly, S100A8 and S100A12 had been already reported to be up-regulated at the transcript level in the work by Bovin *et al.*[45]. Liao and co-workers confirmed that CRP and calgranulin A, B and C were elevated in the sera from erosive RA patients when compared to non-erosive RA patients or to healthy donors, and suggest further validation for their use as markers of disease severity.

Finally, the group of Kantor *et al.* [52] is developing a bioanalytical platform for phenotypic analysis and biomarker discovery in whole blood and other samples, which combines several methods for quantification of proteins (proteome), low molecular weight biomolecules (metabolome), and other several hundred cellular parameters. Their goal is to identify candidate biomarkers and subsequently, to establish measurement assays for these specific markers, which can be performed more easily in a clinical laboratory. They have carried out some initial analysis with samples from RA patients, in order to demonstrate the potential of their methodology. In a near future, we will probably hear more about this promising platform.

From Immunomic Approaches

Two interesting works on the characterization of auto-antibody profiles with micro-array technology should be mentioned here.

The first one is the work by Robinson *et al.* [17]. Here, they describe the development of an auto-antigen array, which included 196 different biomolecules representing major auto-antigens in eight distinct human autoimmune diseases, such as proteins, peptides, enzyme and ribonucleotide complexes, DNA, and post-translationally modified antigens. This micro-array was used for the detection of antibodies in the serum of patients with RA, SLE, polymyositis (PM), primary biliary cirrhosis, Sjögren syndrome, diffuse and limited systemic sclerosis (SSc), and mixed-connective tissue disease. Bound antibodies were detected using anti-human IgM/IgG conjugated with Cy3. The results correlated with those obtained for the same serum samples using conventional methods, such as ELISA, immuno-precipitation and western-blot analysis. Furthermore, they were able to make comparative quantification using internal standards.

The second report by Feng *et al.*[53] is a smaller version of the first one. Here, they selected only 15 auto-antigens, and the detection of bound auto--antibodies was done with horseradish peroxidase--conjugated secondary antibodies and chemoluminiscent substrates. Antibodies were quantified using calibration curves, and positive samples were confirmed by commercially available methods. They analyzed cohorts of patients with RA, SLE, SSc, PM and Sjögren syndrome.

Both groups are confident that their methods represent a powerful high-throughput, low-cost tool for studying auto-antibody responses.

And What do We Get From Genomic Approaches? Perspectives of Genomics in Rheumatic Diseases

Before the breakthrough of genome-wide approaches, most expression studies have evaluated genes or proteins either one by one or a maximal of several dozens at a time[54-58]. The main attractive of these new discovery approaches is the possibility of analyzing several hundred to thousand target genes or proteins at the same time, offering the opportunity to get a global picture of a given pathology [59;60]. Then, one or several interesting candidates, which could be directly involved in the disease process, can be identified and selected for further characterization. As was mentioned before,

this selection depends strongly on the special interests of the investigation group. In any case, examples for "candidates" worthy of further investigation could be:

– a SNP associated to a disease, as for example the possible association between a polymorphism in the human CD19 gene and SLE[40] or in the IL1 gene cluster and RA[43]

– an up or down regulated gene transcript, as was reported for clusterin and for chitinase 3-like 2 gene in the osteoarthritic cartilage[37;39;61]

– an increased or decreased protein level, as in the case of calgranulin A and B as markers for differentiation of RA from other inflammatory joint diseases[62]

– detected auto-antibodies, as described by Feng *et al.* using antigens characteristic for rheumatic diseases[53]

Furthermore, for complex and multi-factorial disorders, such as rheumatic diseases, the search for a single altered gene can be the wrong strategy. Instead, it could be more appropriate to define profiles of expression of several genes or proteins in relation to the diseases, so-called "signatures"[38;47;63].

After identification of an interesting candidate resulting from genome wide approaches, the investigator must decide what to do next with the molecule of interest, as would be explained next. An overview about implications of genome-wide approaches and definition of possible candidate genes or molecules for the clinical practice of rheumatic diseases is given next and is also schematically presented in Figure 2.

Validation

As was mentioned before under "Limitations and possibilities…", validation or confirmation of results from genomic approaches is extremely important. Nevertheless, from all resulting potential candidates, only few of them can be validated and further analyzed, due to the time consuming and extensive experiments required for this. Therefore, the choice of targets for validation should be done based on bioinformatic data, and should be supported by available data regarding their relation to the disease under investigation.

Results from single nucleotide mutations and polymorphisms analysis can be easily confirmed by PCR approaches[64], but there are also reports with other methods, for example with oligonucleotide arrays[65]. Also for validation of candidate genes after transcriptome analysis with microarrays, real time PCR is a preferred validation method nowadays, due to its sensitivity and feasibility, but other methods of choice could be *in situ* hybridization or northern blot approaches[66]. In turn, validation of target proteins after 2-DE and mass spectrometry (MS) approaches can be done for example with immunoassays and multiple reaction monitoring-MS (MRM-MS)[24;67].

Beyond simple replication of experiments with the same samples, validation experiments should consider two aspects:

– candidates or targets should be determined with other methods in the same samples or patients used for the original approach.

– targets should be determined in a larger number of samples or patients with these other methods.

Further Clinical and Laboratory Investigations

Successfully accomplished validation experiments for candidates of interest deliver the basis to decide which of them should be exhaustive investigated. Since many genes and their proteins as well as many molecular networks and pathways are still not yet sufficiently characterized, elucidation and interpretation of functional role in a normal physiological or a pathological process represent the next hurdle in the discovery process. Since a thoroughly characterization of a gene or a protein can in some cases take several months or even years into account, clinical studies for confirmation of positive or negative correlation of the candidate with a disease or a disease phase is done sometimes based only on clinical aspects, but without profound knowledge of its molecular biology and effects.

In this way, the gene or protein, or one of their biological effects can be measured in different cohorts, e.g. patients before and after a treatment, patients with other pathologies, healthy donors, etc. With the results from these studies, possible correlations should be established with several factors, as for example familiar occurrence, risk and other predisposing factors, clinical and laboratory parameters of disease activity, and clinical outcomes.

On the other side, functional confirmation of gene expression profiling can be done with additional arrays for proteins, auto-antigens, auto-antibodies, metabolites or glycoproteins specific for defined cell or tissue types. Nevertheless, for a de-

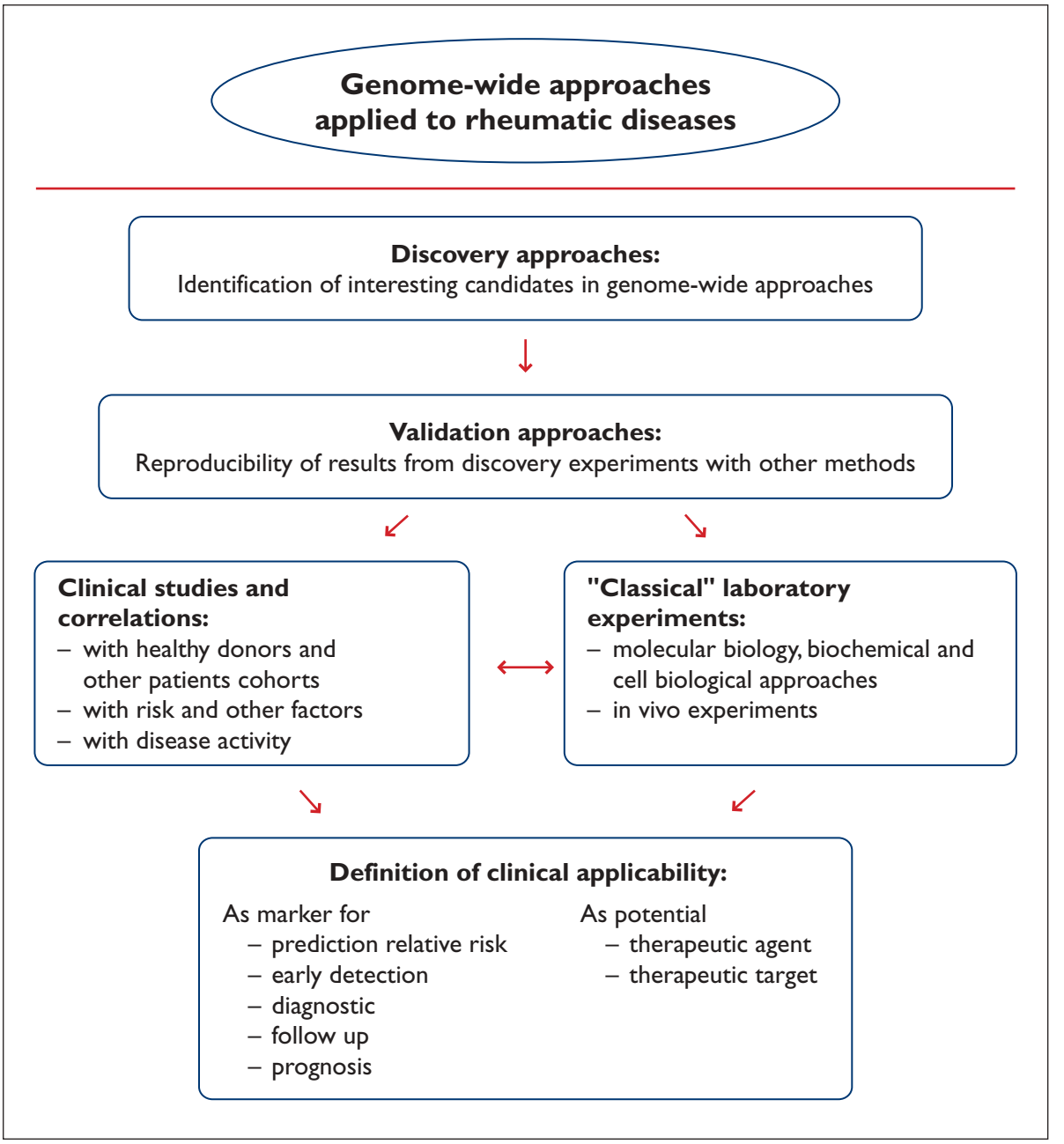

Figure 2. Illustration of the work flow from genome-wide approaches to the definition of clinical applicability of new markers in rheumatic diseases

tailed characterization of single genes or proteins, it is indispensable to go back to classical laboratory work and apply traditional approaches from genetic, biochemistry, and cell biology, in order to gain accurate information on expression, regulation and biological functions of the candidate in question.

Potential Clinical Applicability

After consideration of all molecular, biochemical and functional analysis data on the one side, and of all clinical and correlation data on the other side, it can be determined if the target has potential clinical applicability. With all data from interdisciplinary work, the fully characterized molecule of interest

can be defined as disease biomarker for prediction of relative risk, early detection, diagnosis, follow up, or prognosis. Alternatively, depending on the biological effects and/or functional role, it can be considered as potential candidate for therapeutic approaches, either as therapeutic agent or as target.

Clinical applicability opens the possibilities for another "omics" discipline: pharmacogenomics, which is defined as the study of genetically caused interpersonal variability in drug responses. Numerous publications have shown that genetic polymorphisms localized within the coding regions for drug-metabolizing enzymes, surface receptors, cell transporters and other drug targets can influence drug efficacy and/or toxicity, and that these polymorphisms and therefore their effects vary between individuals. By studying these inter-personal differences at the genetic level, it should be possible to adjust drug prescription based on individual genotypes[68-70]. This would be the ultimate from-bench-to-bedside implication of genome-wide approaches for clinical practice.

Closing Remarks

It has been a long way since 1985, when scientists began to consider the realization of the ambitious project of sequencing the complete human genome until official announcement of the completion of a working draft of the human genome sequence in June 2000. However, this was not the end but just a milestone in the so-called genomic revolution. At that point, many people thought that the nucleotide sequence, as ultimate physical map of the human genome, would be a kind of open book from which complexities of gene expression and genetic mechanisms of disease could simply be read. Nevertheless, it turn out to be not that easy: there were huge expectations of rapid acquisition of knowledge, but it has happen slowly than expected. Considering the above mentioned and other published reports on genome-wide approaches regarding different aspects of Rheumatology, it is evident that until now, only a small part of the work has been done in this area, in a rather fragmentary way.

Now, we are in front of a broad range of sometimes frustrating inconvenients but also exciting possibilities, which all together should help us to understand genetic and molecular mechanisms ruling biological and physiological processes, and to treat effectively or at least, to manage appropriately complex pathological disorders, such as rheumatic diseases. Genome-wide approaches, with all their sub-disciplines, can contribute in this process. In order to accelerate discover, and to optimize personal and technical resources, cooperation and exchange of information between groups needs to be strongly encouraged. An ideal situation would be accessibility to all obtained information from genome-wide approaches at al levels (DNA, RNA, proteins and beyond) in public databases, preferably managed by international consortiums and ruled according to principles of good scientific practice, giving all interested groups worldwide the possibility to participate in this fascinating discovery process.

Acknowledgements

The authors want to thank Dr. Thomas Häupl for critical reading of the manuscript and for the stimulating discussion afterwards.

References

1. Yelin EH, Callahan LF. The economic cost and social and psychological impact of musculoskeletal conditions. Arthritis Rheum 1995;38:1351-1362.
2. Wikipedia. The Free Encyclopedia. Word Wide Web. http://en.wikipedia.org/wiki/-omics. Internet . 2005.
3. McKusick VA, Ruddle FH. A new discipline, a new name, a new journal. Genomics 1987;1:1-2.
4. Harley JB, Trent J, Kastiner DL. American College of Rheumatology Basic Research Conference: Genetics and genomics in rheumatic disease. Arthritis Rheum 2002;47:93-98.
5. Salfeld JP. World leaders heap praise on human genome landmark. Nature 2000;405:983-984.
6. Marshall E. Human genome: Rival genome sequencers celebrate a milestone together. Science 2000;288: 2294-2295.
7. Nature Special Edition: The human genome project. Nature 2001;409:813-958.
8. Science Special Edition: Human Genome. Science 291(5507). 2001.
9. Hieter P, Boguski M. Functional genomics: It's all how you read it. Science 1997;278:601-602.
10. SNP attack on complex traits. Editorial. Nature Genetics 2005;20:217-218.
11. Chakravarti A. Population genetics—making sense out of sequence. Nat Genet 1999;21:56-60.
12. Häupl T, Burmester G-R, Stuhlmüller B. Neue Aspekte der molekularbiologischen Diagnostik. Array-Technologie und Expressionsprofile für die Charakterisierung rheumatischer Erkrankungen. Z Rheumatol 2002;61:396-404.
13. Link AJ, Eng J, Schieltz DM, et al. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol 1999;17:676-682.

14. Adkins JN, Varnum SM, Auberry KJ, et al. Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. Mol Cell Proteomics 2002;1:947-955.
15. Patterson SD. Proteomics: Beginning to realize its promise? Arthritis Rheum 2004;50:3741-3744.
16. Schonbach C. From immunogenetics to immunomics: funtional prospecting of genes ans transcripts. Novartis Found Symp 2003;254:177-188.
17. Robinson WH, DiGennaro C, Hueber W, Haab BB. Autoantigen microarrays for multiplex characterization of autoantibody responses. Nat Med 2002;8:295--301.
18. De Ceuninck F, Dessencourt L, Anract P. The inflammatory side of human chondrocytes unveiled by antibody microarrays. Biochem Biophys Res Communi 2004;323:960-969.
19. Tirumalai RS, Chan KC, Prieto DH, Issaq HJ, Conrads TP, Veenstra TD. Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics 2003;2:1096-1103.
20. Schmidt C. Metabolomics takes its place as latest up-and-coming "omic" science. J Natl Cancer Inst 2005;96:732-734.
21. Dettmer K, Hammock BD. Metabolomics - A new exciting field within the "omics" sciences. Environ Health Perspect 2004;112:396-397.
22. Fiehn O. Metabolomics-the link between genotypes and phenotypes. Plant Mol Biol 2002;48:155-171.
23. Brusic V, Petrovsky N. Immunoinformatics: The new kid in town. Novartis Found Symp 2003;254:3-13.
24. Liao H, Wu J, Kuhn E, et al. Use of mass spectrometry to identify protein biomarkers of disease severity in the synovial fluid and serum of patients with rheumatoid arthritis. Arthritis Rheum 2004;50:3792-3803.
25. Kennedy GC. The impact of genomics on therapeutic drug development. EXS 2000;89:1-10.
26. Häupl T, Krenn V, Stuhlmüller B, Radbruch A, Burmester G-R. Perspectives and limitations of gene expression profiling in rheumatology: new molecular strategies. Arthritis Res Ther 2004;6:140-146.
27. Hagman M. Human genome: A good SNP may be hard to find. Science 1999;285:21-22.
28. Halushka MK, Fan JB, Bentley K, et al. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 1999;22: 239-247.
29. Waldron-Lynch F, Adams C, Amos C, et al. Tumor necrosis factor 5' promoter single nucleotide polymorphisms influence susceptibility to rheumatoid arthritis (RA) in immunogenetically defined multiplex RA families. Genes Immun 2001;2:82-87.
30. Kaijzel EL, Bayley JP, van Krugten MV, et al. Allele-specific quantification of tumor necrosis factor alpha (TNF) transcription and the role of promoter polymorphisms in rheumatoid arthritis patients and healthy individuals. Genes Immun 2001;2:135-144.
31. Bridges SLJr, Jenq G, Moran M, et al. Single-nucleotide polymorphisms in tumor necrosis factor receptor genes: definition of novel haplotypes and racial/ethnic differences. Arthritis Rheum 2002;46: 2045-2050.
32. Shinohara Y, Ezura Y, Iwasaki H, et al. Three TNFalpha single nucleotide polymorphysms in the Japanese population. Ann Hum Biol 2002;29:579-583.
33. Barton AP, Platt H, Salway F, et al. Polymorphisms in the tumour necrosis factor gene are not associated with severity of inflammatory polyarthritis. Ann Rheum Dis 2004;63:280-284.
34. Tsuchiya N, Komata T, Matsushita M, et al. New single nucleotide polymorphisms in the coding region of human TNFR2: association with systemic lupus erythematosus. Genes Immun 2000;1:501-503.
35. Nishibu A, Oyana N, Nakamura K, et al. Lack of association of TNF-238A and -308A in Japanese patients with psoriasis vulgaris, psoriatic arthritis and generalized pustular psoriasis. Dermatol Sci 2002;29:181--184.
36. Bayley JP, Ottenhoff TH, Verweij CL. Is there a future for TNF promoter polymorphisms? Genes Immun 2004;5:315-329.
37. Camargo JF, Correa PA, Castiblanco J, Anaya JM. Interleukin-1beta polymorphisms in Colombian patients with autoimmune rheumatic diseases. Genes Immun 2004;5:609-614.
38. D'Alfonso S, Barizzone N, Giordano M, et al. Two single-nucleotide polymorphisms in the 5' and 3' ends of the osteopontin gene contribute to susceptivility to systemic lupus erythematosus. Arthritis Rheum 2005;52:539-547.
39. Kuroki K, Tsuchiya N, Tsao BP, Grossman JM, Fukazawa T, Hagiwara K. Polymorphisms of human CD19 gene: possible association with susceptibillity to systemic lupus erythematosus in Japanese. Genes Immun 2002;3:S21-S23.
40. Kawaguchi Y, Tochimoto A, Ichikawa N, et al. Association of IL1A gene polymorphisms with susceptibility to and severity of systemic sclerosis in the Japanese population. Arthritis Rheum 2003;48:186-192.
41. Zhou X, Tan FK, Reveille JD, et al. Association of novel polymorphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. Arthritis Rheum 2002;46:2990-2999.
42. Kodera T, Tan FK, Sasaki T, Arnett FC, Bona CA. Association of 5'-untranslated region of the Fibrillin-1 gene with Japanese scleroderma. Gene 2002;297:61-67.
43. Tan FK, Wang N, Kuwana M, et al. Association of fibrillin 1 single-nucleotide polymorphism haplotypes with systemic sclerosis in Choctaw and Japanese populations. Arthritis Rheum 2001;44:893-901.
44. Heller RA, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc Nat Acad Sci USA 1997;94:2150-2155.
45. Bovin LF, Rieneck K, Workman C, Nielsen H, Sørensen SF. Blood cell gene expression profiling in rheumatoid arthritis. Discriminative genes and effect of rheumatoid factor. Immunol Lett 2004;93:217-226.

46. Olsen N, Sokka T, Seehorn CL, et al. A gene expression signature for recent onset rheumatoid arthritis in peripheral blood mononuclear cells. Ann Rheum Dis 2004;63:1387-1392.
47. Dotzlaw H, Schulz M, Eggert M, et al. A pattern of protein expression in peripheral blood mononuclear cells distinguishes rheumatoid arthritis patients from healthy individuals. Biochin Biophys Acta 2004;1696: 121-129.
48. Blass S, Engel GB, Burmester GR. The immunologic homunculus in rheumatoid arthritis. A new viewpoint of immunopathogenesis in rheumatoid arthritis and therapeutic consequences. Z Rheumatol 2001;60:1-16.
49. Liu X, Piela-Schmit TH. Fibrin(ogen)-induced expresion of ICAM-1 and chemokines in human synovial fibroblasts. J Immunol 2000;165:5255-5261.
50. Matcht LM, Elson Cj, Kirwan JR, et al. Relationship between disease severity and responses by blood mononuclear cells from patients with rheumatoid arthritis to human heat-shock protein 60. Immunology 2000;99:208-214.
51. van Room JA, van Eden W, van Roy FJ, et al. Stimulation of suppressive T cell responses by human but not bacterial 60-kD heat-shock protein in synovial fluid of patients with rheumatoid arthritis. J Clin Invest 1997;100:459-463.
52. Kantor AB, Wang W, Lin H, et al. Biomarker discovery by comprehensive phenotyping for autoimmune diseases. Clin Immunol 2004;111:186-195.
53. Feng Y, Ke X, Ma R, Chen Y, Hu G, LIu F. Parallel detection of autoantibodies with microarrays in rheumatoid diseases. Clin Chem 2004;50:416-422.
54. Stuhlmüller B, Kunisch E, Franh J, et al. Detection of oncofetal h19 RNA in rheumatoid arthritis synovial tissue. Am J Pathol 2003;163:2645.
55. Egerer K, Hertzer J, Feist E, et al. sE-selectin for stratifying outcome in rheumatoid arthritis. Arthritis Rheum 2003;49:546-548.
56. Tchetverkov I, Lard LR, DeGroot J, et al. Matrix metalloproteinases -3, -8, -9 as markers of disease activity and joint damage progression in early rheumatoid arthritis. Ann Rheum Dis 2003;62:1094-1099.
57. Stuhlmüller B, Ungethüm U, Scholze S, et al. Identification of known and novel genes in activated monocytes from patients with rheumatoid arthritis. Arthritis Rheum 2000;43:775-790.
58. Seki T, Selby J, Haupl T, Winchester R. Use of differential subtraction method to identify genes that characterize the phenotype of cultured rheumatoid arthritis synoviocytes. Arthritis Rheum 1998;41:1356-1364.
59. Lequerre T, Coulouarm C, Derambure C, et al. A new tool for rheumatology: large-scale analysis of gene expression. Joint Bone Spine 2003;70:248-256.
60. Elferink, R.O. One stepp further towards real high-throughput functional genomics. Trends Biotechnol 2003;21.
61. Knorr T, Obermayr F, Bartnik E, Zien A, Aigner T. YKL--39 (chitinase 3-like protein 2), but not YKL-40 (chitinase 3-like protein 1), is up regulated in osteoarthritic chondrocytes. Ann Rheum Dis 2003;62:995-998.
62. Drynda S, Ringel B, Kekow C, et al. Proteome analysis reveals disease-associated marker proteins to differentiate RA patients from other inflammatory joint diseases with the potential to monitor anti-TNF alpha therapy. Pathol Res Pract 2004;200:165-171.
63. Han GM, Chen SL, Shen N, Ye S, Bao CD, Gu YY. Analysis of gene expression profiles in human systemic lupus erythematosus using oligonucleotide microarray. Genes Immun 2003;4:175-176.
64. Johnson MP, Haupt LM, Griffiths LR. Locked nucleic acid (LNA) single nucleotide polymorphism (SNP) genotype analysis and validation using real-time PCR. Nucleic Acids Res 2004;32:55.
65. Dong S, Wang E, Hsie L, Cao Y, Chen X, Gingeras TR. Flexible use of high-density oligonucleotide arrays for single-nucleotide polymorphism discovery and validation. Genome Res 2001;11:1418-1424.
66. Blackhall FH, Wigle DA, Jurisica I, et al. Validating the prognostic value of marker genes derived from a non-small cell lung cancer microarray study. Lung Cancer 2004;46:197-204.
67. Kuhn E, Wu J, Karl J, et al. Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards. Proteomics 2004;4:1175-1186.
68. Evans WE, Relling MV. Pharmacogenomics: Translating functional genomics into rational therapeutics. Science 1999;286:487-491.
69. Destenaves B, Thomas F. New advances in pharmacogenomics. Curr Opin Chem Biol 2000;4:440-444.
70. Vesell ES. Advances in pharmacogenetics and pharmacogenomics. J Clin Pharmacol 2000;40:930-938.

Address for correspondence:
Dipl.-Biol. Lorena Martinez Gamboa
Charite University Hospital – Campus Mitte
Medical Clinic III for Rheumatology, Clinical Immunology and Osteology
Laboratory # 329 at the German Center for Investigations on Rheumatic Diseases (Deutsches Rheuma-Forschungszentrum Berlin, DRFZ)
Schumannstraße 21/22
10117 Berlin, Germany
Telefon: +49-30-28460-665
E-mail: martinez-gamboa@drfz.de
lorena.martinez_g@charite.de