Abstract

Osteoarthritis (OA) is one of the most prevalent rheumatic diseases affecting more than half of the population above sixty years of age. The underlying disease processes are slowly being uncovered, but research is often focussed on a single disease process (e.g. collagen degradation or inflammation), a specific therapeutic target (e.g. aggrecanase-2) or risk factor (e.g. association to a certain gene). Drug development projects addressing single therapeutic targets have been largely unsuccessful in OA, as were approaches to describe the OA disease process using single biomarkers. As such, understanding and eventually treating this complex, multifactorial chronic disease, appears to require a different approach that takes into account the myriad of disease processes and complex biology.

Systems biology aims to describe and to understand complex biological systems with the goal of developing predictive models of human disease. To this purpose systems biology approaches integrate information from large datasets of ‘omics’ data (transcriptomics, proteomics & metabolomics) derived from complex in vitro (cell & tissue culture) and in vivo (animal & patients) studies into an in silico model of the disease. This way, systems biology provides a framework for understanding biological processes in health and disease and contributes to the identification of therapeutic targets.

In this paper, the published ‘omics’ data for osteoarthritis are reviewed from a systems biology (hence data integration) point of view; gaps in the available data that prevent full-fledged systems biology evaluation of OA are pinpointed, as well as lessons learned from the studies that have been conducted.

**Systems biology: concept, data sources and work flow**

Systems biology can be defined as ‘studying biology as an integrated system of genetic, protein, metabolite, cellular, and pathway events that are in flux and interdependent’ (1). While such a holistic, systems-based approach emerged across various scientific domains in the last century, it only recently gained considerable momentum in the biological sciences and pharmaceutical research as a result of advances in analytical technologies and computational techniques, which culminated in the ‘omics’ technologies (2;3). These ‘omics’ technologies enable the simultaneous measurement of hundreds or even more molecules in a single analysis, thereby providing the building blocks (raw data) from which biological interaction networks can be reconstructed, making a systems approach feasible (4).

The ‘omics’ technologies are developed for the qualitative and quantitative analysis of the major classes of biological relevant molecules: genomics for DNA, transcriptomics for RNA, proteomics for proteins, and metabolomics for metabolites. While genomics technologies like genome-wide linkage approaches are powerful with regard to finding disease-associated patterns in the DNA, our focus is rather on the systems behaviour of the gene expression products (transcriptomics and proteomics) and metabolites (metabolomics). A widely used technology in transcriptomics research is the cDNA array, allowing the profiling of gene expression by their RNA transcription products (5;6). For proteomics experiments two-dimensional gel electrophoresis (2-DE) is still the most widely used technique for OA research (7-11). After 2-DE of protein extracts from ‘diseased or treated’ and control tissue, proteins whose abundance varies between the two groups, are excised from the gel and *in-situ* digested by a protease. Identification of the resulting peptide mixture is then performed by mass spectrometry (MS)(12). Alternatively, peptides (derived from digested protein mixtures) can be analyzed by directly single –or multidimensional liquid chromatography coupled to MS (LC-MS). Metabolomics is the youngest of the three technologies whose methods are still subject to further development. Compared to the genome and proteome, however, the metabolome is closest to the actual organism’s phenotype and therefore also closely resembles the various disease states. In addition, among various species the metabolic
pathways are conserved and the majority of the metabolites are the same (whereas gene and protein sequences may vary), making the process of translation between species easier. Therefore, it is expected that metabolomics will contribute substantially to our understanding of disease development and progression.

As metabolomics needs to deal with an extreme diversity in chemical structures and concentration range (13-15), in general multiple analytical approaches are required to cover the broad range of the metabolome. Metabolomics can be divided into more global screening platforms based on for example nuclear magnetic resonance (NMR) or gas chromatography-mass spectrometry (GC-MS), and (often liquid chromatography-MS(LC-MS)-based) platforms targeted to specific types of chemical classes, i.e. amino acids, lipids, or endogenous peptides.

The typical workflow of a systems biology experiment is shown in Figure 1. Given a biological question/hypothesis derived from existing knowledge, an experimental plan/design is set up, covering all necessary decisions for the experiment (which and how many samples, what technical platform; such as untargeted, targeted, flux analysis, etc). Subsequently, the selected samples are treated to extract the RNA transcripts, proteins, and/or metabolites, followed by their measurement (untargeted or targeted) with the respective analytical methods. The resulting raw data need then to be processed with various statistical and informatics tools to extract the relevant information, finally resulting in a systems model specific for the sample type and group of compounds under study, which can be biologically interpreted. This model is then incorporated into a master model that unifies information from multiple organisms, tissues and compartments (SF, blood) and in addition time points. This enables a comprehensive interpretation of the causality of the involved mechanisms and translatability of the used animal models. The addition of the new information will address posed questions/hypotheses and trigger new ones that need to be experimentally verified, thus creating a reiterative process.
It should be emphasized that the best possible comprehensive view on the biological system can only be acquired by combining the information from the genomics, transcriptomics, proteomics, and metabolomics level for both biological and analytical reasons. It has been shown that there is generally a low concordance between RNA expression and protein expression, which is in part due to regulatory mechanisms involved in the translation of the transcripts and post-translational modifications on the protein, resulting in non-linearity. A good example is a study where only 28% of the differentially expressed proteins coincided with RNA differential expression (16). In addition, each analytical method inherently has a bias towards certain molecules, making the resulting biological ‘picture’ incomplete.

An example is the NMR metabolomics platform, which is biased towards abundant metabolites, often represented by molecules related to energy metabolism. By combining methods, this bias is partially alleviated and a more complete biological ‘picture’ emerges.
Figure 1. The envisioned workflow of a systems biology experiment. Given a biological question/hypothesis derived from existing knowledge and earlier experiments, the appropriate experimental design is determined (including sample type (human, in vivo, or in vitro) and sampling protocol). The samples are then analyzed for their transcript, protein, and metabolite content, followed by the extraction of information from the raw data, statistical evaluation, and biological interpretation. This newly acquired biological information is then integrated into a master model that unifies information from multiple tissue types, compartments, organisms, and time points and will likely spark new questions and hypotheses. New experiments are then performed to validate new hypotheses, effectively creating a reiterative circle of hypothesis creation and validation.
Osteoarthritis (OA) is one of the most prevalent rheumatic diseases affecting more than half of the population above sixty years of age (17). The main characteristic of the disease is progressive loss of articular cartilage which is thought to originate from an imbalance between synthesis and degradation of the cartilage matrix (18;19). However, its precise aetiology is still far from understood and recent insights are pointing towards the increasingly supported view that OA is a disease resulting from multiple pathophysiological mechanisms in which local and systemic factors, as well as biomechanical triggers are interplaying (17). Due to our limited understanding, no disease modifying treatments are currently available, and therefore the only existing therapies primarily comprise analgesics. Furthermore, there is a lack of adequate biomarkers for early diagnosis, prediction of eventual joint damage, assessment of disease activity and monitoring of the efficacy of therapies (20-22).

As OA pathology likely comprises a multitude of systemic & local factors and pathophysiological mechanisms (see Figure 2), a more holistic methodology, complementary to the classical biological approach of reductionism represents a valuable addition for finding clues about the disease aetiology and progression. To do so, transcriptomics, proteomics and metabolomics data sets should be generated from (pre)clinical studies. As we are most interested in an OA systems biology model relevant for humans, it would be logical to use samples from human origin. However, well-defined human OA-relevant samples from the affected joints (synovial fluid, synovial tissue, ligaments and cartilage) are scarce, making it often necessary to use samples from other species. It is to be expected that systems biology models derived from these non-human samples only partially resemble the human situation as is aptly exemplified in a study where changes detected in protein expression were due to chondrocyte dedifferentiation rather than to disease processes (23). Therefore, much attention needs to be directed towards translational research, in order to better understand the commonalities and differences between the various disease models at a systems level.
The goal of this study was to integrate the data from the available ‘omics’ publications related to OA, so as to come to a first, systems view of the disease. As the data sources are still relatively scarce, this effort should be seen as an explorative effort to assess the current situation and to identify knowledge gaps that currently prevent an integrated view on OA.

We found 24 research papers employing transcriptomics, proteomics, and/or metabolomics approaches. These studies all addressed a single level of molecules (mRNA, protein or metabolites) and none of the studies combined their data into a systems biology view of the disease. The samples studied ranged from *in vitro* chondrocyte cultures to human knee synovial fluid and guinea pig urine samples. We integrated the data from these studies to find common (occurring in multiple articles) clues for new disease mechanisms and also the potential causality of these mechanisms by integrating data obtained in different compartments, to such an extend as possible by the published data.
When data from these studies are combined, it is clear that aberrant cartilage homeostasis is involved in the OA disease, which is in accordance with contemporary views arisen from molecular biological and biochemical studies. However, some of the papers also revealed a number of potentially important but less studied mechanisms (see supplement Table 1 for details). The three likely most relevant mechanisms, that have been reported in multiple studies and on multiple levels (transcript, protein, and/or metabolite) will be discussed in more detail below: carbohydrate & lipid metabolism, cell signalling, and oxidative stress defense.

**Carbohydrate & lipid metabolism.** Many of the ‘omics’ based studies in OA reveal alterations in molecules that are involved in carbohydrate and lipid metabolism (see Figure 3). This is especially true for the metabolites and proteins in relation to OA. NMR-based metabolomics studies in canine SF and human and guinea pig urine (24-27), report strikingly consistent results despite the fact that these sample types are quite diverse with respect to species and biological fluid. The general picture that emerges from these experiments is an impaired aerobic glycolytic metabolism in favour of increasing lipolysis. In SF this process causes a hypoglycaemic, hypoxic and acidic environment; lower glucose concentrations were observed whereas lactate and pyruvate increased significantly, being a hallmark of anaerobic metabolism. Changed levels were observed for alanine which is closely linked to pyruvate, malic acid, and hypoxanthine which are involved in purine metabolism. Interestingly, no abnormal readings of citrate were reported in any of the papers. This suggests that an altered activity of the citric acid cycle is not the cause of diminishing aerobic energy production. Increased lipolysis is indicated by elevated concentrations of glycerol, being a product of triglyceride hydrolysis, and by increased levels of ketone bodies, such as hydroxybutyrate. It is postulated that hydroxybutyrate and other ketone bodies act in a regulatory mechanism that maintains glycolytic intermediates within certain boundaries for biosynthetic purposes. Furthermore, increased lipoprotein concentrations in SF were reported (25;26).
Although the NMR-based metabolomics studies indicate an important role for the energy system in osteoarthritis, it should be mentioned that this analytical method has a bias towards the highly abundant molecules that play a role in the energy system. Therefore, it is possible that its role is overemphasized, making it necessary to corroborate these findings with metabolomics data using other analytical platforms and levels of biological information in a true systems biology approach. For the mechanisms described above, also at the protein level evidence is found. In a proteomics experiment with cultured human chondrocytes on alginate beads, various energy system related proteins were found to be differentially expressed (28). Proteins involved in the conversion of glucose into pyruvate (alpha-enolase and triose-phosphate isomerase) were found to be upregulated as were proteins involved in the pentose-phosphate pathway (transaldolase and 6-phospho-gluconolactonase) and oxidative phosphorylation (ATP-synthase (α-subunit)).
In similar fashion, another proteomics study described higher levels of aldehyde dehydrogenase 3 [NAD(P)], glucosidase II α subunit, and phosphoglucomutase 1, compared to the control samples (29). However, contradictory to the previously discussed article (although both studied cultured human chondrocytes) they found lower concentrations of alpha-enolase, glyceraldehyde 3-phosphate and fructose biphosphate aldolase. In a proteomics study using non-cultured cartilage samples, again higher levels were found for alcohol alpha-enolase, pyruvate kinase 3 isoform 2 (involved in the production of pyruvate), and alcohol dehydrogenase, whereas reduced concentrations were reported for flavin reductase and adenylate kinase isoenzyme 1 (30). In a separate study comparing mesenchymal stem cells from OA and control patients, also similar pronounced changes in metabolism were observed. These bone marrow derived cells exhibit a multipotent differentiation potential and are known to migrate into damaged zones and to be involved in tissue remodelling (31).

In two separate studies in which also extracellular proteins were analyzed, apolipoproteins were found to be higher in OA (32;33). In the first study non-cultured human chondrocytes from seven OA patients and seven healthy donors were compared, and it was found that Apo A-I, II, IV, and Apo H were significantly elevated in the osteoarthritis samples (32). In the second experiment SF samples from control individuals (n = 20), patients with early osteoarthritis (n = 21), and patients with late osteoarthritis (n = 21) were compared and here, in addition to apolipoprotein H, also apolipoprotein E was found to be elevated (33). Apolipoproteins are the protein elements in the lipoprotein particles and fulfil several lipid-related tasks. Due to the multitude of their functions, it is difficult to interpret the biological relevance of their upregulation. It could be, for instance, that their upregulation is indeed related to the increased lipid metabolism reported earlier, or that it is a consequence of inflammatory processes. It could also be that the apolipoproteins are released as a consequence of cartilage degradation.

The key question relating to the changes in the energy system is whether it is a cause or a consequence of the disease. It could be that these changes are the mere effect of the activation of the chondrocytes, in an attempt to repair the cartilage damage. But reports of increased incidences of hand
osteoarthritis in obese patients, suggest a systemic biochemical relation (34) and gives rise to the hypothesis that osteoarthritis is a systemic disorder in which altered energy metabolism in general and lipid metabolism in particular may predispose to OA development.

While metabolomics approaches have mapped carbohydrate metabolism-related changes in OA, no method has been applied to directly and comprehensively study OA-related changes in the lipid profiles. More insight into the role of lipid metabolism in osteoarthritis could therefore be acquired by a sensitive analysis platform that allows the system-wide analysis of lipids, as for example a recently published LC-MS method (35). It would be especially interesting to analyze lipid species involved in energy metabolism and signalling (cholesterol esters, sphingomyelins, triglycerides, etc.) in early osteoarthritis samples to test the hypothesis that lipids play a crucial role in OA.

**Cell signaling.** While inflammatory processes are strongly implicated in rheumatoid arthritis, their involvement in osteoarthritis is less evident. However, also for OA, both on the transcriptomics and proteomics level, differential regulation of inflammatory mediators has been observed. Transcriptomics experiments conducted on cartilage samples primarily picked up changes in the expression of various cytokines. Surprisingly, in one study down regulation was reported of many genes involved in the interleukin-1 (IL-1) pathway, including IL1B, IL-6, IL-8, leukaemia inhibitory factor (LIF), and functional antagonist/scavenger receptor type II (36). However, the authors stress that these results still need to be confirmed. In another study it was shown that a reasonably predictive classification model for mild OA versus control could be generated using a blood leukocyte gene expression profiling approach (37). A linear combination of nine genes was found to give a sensitivity of 72% and a specificity of 66% on a blind data set. Amongst these genes were interleukin-13 receptor α-1 (IL13RA1) which is critically important for the IL-13 expression in response to inflammation, and tumour necrosis factor-α-induced protein 6 (TNF-AIP6), which is suspected to have a chondrocyte protective role. In two studies comparing the gene expression of intact and damaged zones of OA cartilage, this gene was also found to have increased in the damaged zones,
while cartilage fibroblast growth factor 13 was found to be downregulated (38;39).

In contrast, none of the proteomics experiments detected cytokines, which is most likely due to their low endogenous concentrations as well as their relatively small size. As various other protein elements important for inflammation are also present only at low concentrations, it is expected that current proteomics technologies will only give a partial and therefore an incomplete view on amongst others the inflammation process.

This shortcoming demonstrates the need for additional platforms focusing on subsections of the proteome. A good example is a method developed to profile the low-molecular-mass proteome, or peptidome, in synovial fluid (40). Comparison of the synovial fluid endogenous peptide profiles of OA, RA, and healthy donors, revealed differences in protease activity and levels of the bioactive peptide bradykinin, which may play an important role in OA (41;42).

**Oxidative stress defense.** Changes in the oxidative cellular defense mechanisms have consistently been found on the protein as well as the RNA level. While the oxidative defense system is reported to be activated in the case of RA, the general trend in OA is towards inhibition. In proteomics analyses peroxiredoxin was upregulated, the results for superoxide dismutase (SOD) Mn were contradictory, and SOD Cu/Zn and thioredoxin-dependent peroxide reductase mitochondrial precursor (PRDX3) were down regulated (28;30). In a gene expression profiling experiment on cartilage samples also an overall decrease in the oxidative damage defense was observed (36). This was specifically the case for glutathione peroxidise 3 (GPX3), SOD 2 and 3, and thioredoxin-interacting protein (TXNIP), of which the down regulation of GPX3 and SOD2 was confirmed with quantitative PCR. In another gene expression study in rat, glutathione reductase was found to have increased concentrations. The role of this relatively unknown process in the pathology of OA is still quite unclear. More insight in the time course and severity of this process could be gained by applying an analytical method aimed at the quantification of oxidative stress.
Future steps

As illustrated by the experiments discussed in the previous section, there is an emphasis in OA research on biochemical changes that occur in the joints. However, significant differences in the expression of various molecules were also observed in bone, urine, and blood. These findings support the view that systemic factors are also involved in OA, and while it is still unclear if these non-local processes are triggers, modulators, or merely a secondary effect, investigating them from a systems-perspective may provide us with new clues and hypotheses with respect to disease aetiology and subsequently potential targets of therapy. This is especially so as elucidating the causality of the mechanisms involved in the various compartments (SF, blood, etc) is one of the strong points of systems biology.

As the data sources we used for this review were scarce and heterogeneous in nature, a full-fledged, integrated systems view of OA is not yet possible. Therefore our work presents an explorative effort that focuses on identifying knowledge gaps so that this systems view can be most efficiently achieved. It should be stressed however, that already this explorative effort reveals the involvement of mechanisms that until now have received relatively little attention, giving a taste of how systems biology can change our view on OA pathology.

Perhaps the most interesting mechanisms are carbohydrate and lipid metabolism. Although the data reviewed in this paper is heterogeneous with respect to sample types and the number of samples studied in the various experiments as well as the analytical methods employed, evidence that energy metabolism changes are associated with OA is surprisingly consistent. As can be seen from Figure 3, quite a significant proportion of molecules at the protein and metabolite level with reported aberrant concentrations in OA have a role in energy metabolism (energy production, carbohydrate metabolism, lipid metabolism). At this point, it is still unclear how these mechanisms relate to each other and whether the process of energy metabolism has a causal relation with the disease or that it is merely resulting from the imbalance in cartilage homeostasis. As evidence for OA associated changes in lipid metabolism exists but many lipid species have
never been actually directly analyzed in relation to OA, their analysis would be particularly informative. Their involvement could be elucidated by using a platform for the system-wide analysis of lipids in plasma samples of patients showing various grades of OA, from very early signs of OA to late stage OA.

As such, a dedicated systems biology approach in which data on the various levels (transcriptome, proteome and metabolome) from various models and species is systematically integrated could provide new insights. While care must be taken to ensure the quality of the experimental design and sampling, as it can significantly influence the results, new insights and advances can be expected in the areas of amongst others translation between disease model species and man, causal mechanisms and progression of the disease over time, and eventually personalized medicine.

Of crucial importance for the success of systems biology is the development of innovative applications for the advanced visualization and interpretation/modelling of newly acquired data, as well as its integration into the already existing body of knowledge. To this end multiple bioinformatics applications have been developed, of which the Ingenuity Pathways analysis program is a good example (43). A central feature of this program is a database extracted from scientific literature via text mining, with information about individual compounds, the role of these compounds in various biological processes, and information on well characterized pathways. After correlation/statistical analysis has been performed on the newly acquired data to determine which compounds show higher or lower levels in disease when compared to a control, this new information can be superimposed onto the already existing knowledge. This way participating biological processes and pathways can be identified and visualized, making a detailed analysis possible, not only in terms of involvement in the disease pathology, but also in terms of the ‘quality’ of the data. For example, this way it can easily be determined if the results from multiple studies (from the same species but different labs/clinics, or from different species) are mutually agreeing with respect to the up- or downregulation of certain molecules or pathways. Also, assuming that the database information is complete, it can be determined if the changes of multiple molecules in a
single pathway make sense when taking into account their influence on each other (i.e. if a molecule has an inhibitory effect on another compound’s synthesis (through interaction with an enzymatic system), does its downregulation result in upregulation of the other compound?), and for what important elements in a pathway expression data is still missing and therefore will need to be measured to validate/exclude the involvement of the pathway.

This is aptly demonstrated by Figure 4, which shows the glycolysis and gluconeogenesis pathway with information about changes in expression for involved compounds, as was found in the reviewed articles above. For example alcohol dehydrogenase (compound 1.1.1.2) was found to be down regulated as a transcript in rat cartilage samples (44), but higher levels were found in OA at the protein level in another study of human knee cartilage (30). Likewise, the lower level in OA of lactate dehydrogenase (1.1.1.27), which consumes lactate to produce pyruvate, should result in higher lactate levels as is observed, but lower levels of pyruvate (24-27;44). Contradictory, also higher levels for pyruvate were observed, most likely due to an increased activity of phosphoenolpyruvate kinase (compounds 2.7.1.40), which was also observed in one of the studies (30).

In conclusion, a holistic view on the disease with information about the patterns of relations and interdependencies of individual pathways, and processes in relevant disease models and compartments (SF, blood, cartilage) as well as in time, in combination with the more targeted approach of fully characterizing changes in (known) individual pathways, holds great promise for better understanding of a complex disease as OA.
Figure 4. Example of a partially resolved, well-known pathway (the glycolysis and gluconeogenesis pathway, green represents reduced concentration and red represents increased concentration). Visualizing the acquired data in this manner makes it possible to see whether the differential expressions of directly related molecules in the pathway are in mutual agreement (i.e. make biochemical sense). In addition, it allows for the identification of elements in the pathway for which the expression remains unresolved (i.e. not measured) but is critical to validating the involvement of the pathway in the disease.
Acknowledgements

The authors would like to acknowledge the financial support from the Centre for Medical Systems Biology (CMSB).

Reference List

(1) J van der Greef; P Stroobant; R van der Heijden. The role of analytical sciences in medical systems biology. *Curr Opin Chem Biol.* 2004, 8, 559-565.
(2) J van der Greef; S Martin; P Juhasz; A Adourian; T Plasterer; ER Verheij; RN McBurney. The art and practice of systems biology in medicine: mapping patterns of relationships. *J Proteome Res.* 2007, 6, 1540-1559.
(4) CB Clish; E Davidov; M Oresic; TN Plasterer; G Lavine; T Londo; M Meys; P Snell; W Stochaj; A Adourian; X Zhang; N Morel; E Neumann; E Verheij; JT Vogels; LM Havekes; N Afeyan; F Regnier; J van der Greef; S Naylor. Integrative biological analysis of the APOE*3-leiden transgenic mouse. *OMICS.* 2004, 8, 3-13.
(16) P Lorenz; P Ruschpler; D Koczan; P Stiehl; HJ Thiesen. From transcriptome to proteome: differentially expressed proteins identified in synovial tissue of patients suffering from rheumatoid arthritis and osteoarthritis by an initial screen with a panel of 791 antibodies. *Proteomics.* 2003, 3, 991-1002.
(31) R Rollin; F Marco; E Camafeita; E Calvo; L Lopez-Duran; JA Jover; JA Lopez; B Fernandez-Gutierrez. Differential proteome of bone marrow mesenchymal stem cells from osteoarthritis patients. *Osteoarthritis Cartilage*. **2008**, 16, 929-935.
(35) C Hu; J van Dommelen; R van der Heijden; G Spijksma; TH Reijmers; M Wang; E Sleec; X Lu; G Xu; J van der Greef; T Hankemeier. RPLC-Ion-Trap-FTMS method for lipid profiling of plasma: method validation and application to p53 mutant mouse model. *J Proteome Res*. **2008**, 7, 4982-4991.


(38) M Geyer; S Grassel; RH Straub; G Schett; R Dinsen; J Grifka; S Gay; E Neumann; U Muller-Ladner. Differential transcriptome analysis of intraarticular lesional vs intact cartilage reveals new candidate genes in osteoarthritis pathophysiology. *Osteoarthritis Cartilage.* **2008**, 17, 328-335.


(40) JJ Kamphorst; R van der Heijden; J DeGroot; FP Lafeber; TH Reijmers; B van El; UR Tjaden; J van der Greef; T Hankemeier. Profiling of endogenous peptides in human synovial fluid by nanoLC-MS: method validation and peptide identification. *J Proteome Res.* **2007**, 6, 4388-4396.

(41) JJ Kamphorst; R van der Heijden; TH Reijmers; J DeGroot; J van der Greef; T Hankemeier. Synovial fluid peptidomics for the study of osteoarthritis. Submitted for publication. **2009**.


(43) Data were analyzed through the use of Ingenuity Pathways Analysis (Ingenuity® Systems, [www.ingenuity.com](http://www.ingenuity.com)). **2009**.