DOI:
10.1038/nrrheum.2017.5

Document Version
Peer reviewed version

Genetics and metabolomics to advance knowledge of rheumatologic diseases

Cristina Menni[1], Jonas Zierer[1], Ana M Valdes[1,2], Tim D Spector[1]

[1] Department of Twin Research and Genetic Epidemiology, Kings College London, London SE1 7EH, UK.

[2] Academic Rheumatology Clinical Sciences Building, Nottingham City Hospital, Hucknall Road, Nottingham, NG5 1PB, UK.

Corresponding author

Prof. Tim D Spector

Department of Twin Research and Genetic Epidemiology, King's College London,

St. Thomas' Hospital, London SE1 7EH, United Kingdom

Tel: +4420 7188 5555; Fax: +44 20 7188 6761

Email: tim.spector@kcl.ac.uk

ABSTRACT

Metabolomics is an exciting field in systems biology which provides a direct signature of the biochemical activities in an individual at a particular time point. Metabolite levels are influenced by many factors including disease status, environment, medications, diet and, importantly, genetics. Their dynamic nature makes them useful for diagnosis or prognosis as well as for predicting and monitoring treatment efficacy. At the same time the strong links to genetic variation allows investigation of pathways underlying changes in their levels. Thus, for metabolomics to yield its full potential, the stable genetic contribution to metabolites and how these link to disease processes needs to be better understood. Here we discuss methodological aspects related to metabolomic profiling. We then review the metabolomic and genetic studies carried out to date on rheumatologic traits and investigate any potential links to the genetics of some of the most common rheumatologic disorders, as well as links to new fields like the gut microbiome.

INTRODUCTION

Metabolomics is a high-throughput technology that captures the global metabolic state of an individual by assaying simultaneously an extensive set of low-molecular-weight compounds (metabolites) in a biological sample[1]. Metabolites can be considered as the intermediate outcome of a number of physiological processes including disease processes, lifestyle, environmental exposure, pharmacological treatments and others (**Figure 1**). In this sense, metabolites are closer to the actual disease phenotype (e.g. rheumatoid arthritis) than genetic variation, gene expression levels or protein levels. This makes metabolomics a powerful tool to identify biomarkers that can be used as diagnosis or to improve prognosis and predict and monitor treatment efficacy, as has been done for systemic autoimmune diseases[2]. Moreover, clinical response to therapy has been shown to result in different metabolite profiles[3] suggesting that metabolomic profiling helps to identify responders to therapy and to improve clinical management of rheumatologic disorders. Metabolomic profiling has therefore the potential to provide an alternative perspective of the altered molecular processes responsible for the onset and pathogenesis of various forms of rheumatologic disease and this can be of paramount importance to prevent joint damage and disability.

At the same time, metabolite levels are also influenced by an individual's genetic makeup[4-9]. In order for metabolomics to yield its full potential the genetic contribution to metabolites and to disease processes needs to be more fully addressed. In this review we cover first the various methodological aspects related to metabolomic profiling. We then review the metabolomic and genetics studies carried to date on selected rheumatologic traits. Finally, we investigate any potential links to the genetics of some of the most common rheumatologic disorders including rheumatoid arthritis (RA), osteoarthritis (OA), psoriasis arthritis (PsA), gout and systemic lupus erythematosus (SLE).

TECHNOLOGICAL STRENGTHS AND LIMITATIONS

Targeted versus untargeted identification

Thousands of small molecules make up the human metabolome[10] and the precise number of independent metabolites potentially detectable is still unclear. Although it is now possible to characterise hundreds of known metabolites, it is not yet possible to get a full coverage of the metabolome. There are two approaches to measure the metabolome: targeted and untargeted (**Figure 2**) [11,12]. Untargeted metabolomics platforms can measure hundreds of metabolites with the advantage of detecting previously unpredicted metabolic perturbations associated with a certain disease. No a priori knowledge on biologically relevant metabolites is required when designing this type of experiments, similar to genome-wide association studies. These approaches are useful to find novel mechanisms or biomarkers. Targeted approaches, on the other hand, resemble candidate gene association study where a smaller and pre-defined set is measured, typically focusing on a few pathways of interest. Advantages include higher sensitivity and absolute (instead of relative) quantification and easy compound identification[11].

Analytical tools

Metabolites vary in size, polarity and concentration making unbiased detection, identification and quantification of the whole metabolome technically challenging[10]. Metabolomics is mainly achieved by either high nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry (MS)[13-15] as recently reviewed[16].

Briefly, NMR spectroscopy is a quantitative non-discriminating, non-destructive, perturbing technique that gives detailed information on solution-state molecular structures, based on atom-centred nuclear interactions. There are several advantages to NMR which include its potential for high-throughput measure, the minimal requirements for sample preparation, reproducibility of the results, the low cost of the measurement, the speed of data acquisition and the non-discriminating

and non-destructive nature of the technique. However, it has relatively low sensitivity and can detect only medium to high abundance metabolites.

On the other hand, MS based metabolomics provides quantitative analyses with high selectivity and sensitivity and the potential to identify metabolites. Samples are initially separated using chromatography (e.g. gas chromatography (GC), capillary electrophoresis (CE) or liquid/ultra performance liquid chromatography (LC/UPLC)) and then identified with a mass spectrometer[17,18] although pre-separation is not always required and there are commercial MS kits (Biocrates) that use direct sampling from plates without pre-separation for quantitative analysis. MS methods provide metabolite separation in a time dimension, from which highly specific chemical information can be extracted. However, MS-based techniques require preparation steps that can be very elaborate and may cause metabolite loss. Also, different platforms can detect different metabolites and therefore, parallel application of techniques, for example, GC-MS and LC-MS is desired to study the metabolome comprehensively.

Sample Type

Metabolomics profiling can be conducted on a variety of biological samples, ranging from urine and blood, to faeces, synovial fluid (SF), saliva, organ tissues, specific cells, tumours, etc.

Urine and blood (serum and plasma) are the most accessible bio-samples, they can be easily and minimally invasively collected in both cases and controls. Also, there is no need for grinding/fractionation. However, they may not reflect biochemical changes in the joint, nor be directly linked to physiologic pathological changes in the joint.

For rheumatologic disorders, synovial fluid and tissues from either the joint or the bone are also very relevant. . SF samples are easy to collect from RA and OA patients with effusion, although less so from unaffected controls. This type of fluid has a lower need for processing and sample preparation than other types of tissue-derived samples (e.g. cartilage, bone). On the other hand, the

metabolomics profile in SF may not fully reflect biochemical changes in other joint tissues such as bone/cartilage/tendons.

Tissues from the joint/bone are very useful when studying pathogenic processes in the joint. Those however are extremely challenging to collect in large numbers particularly for controls. Moreover, LC/MS often requires manual extraction, solvent extraction of tissue samples precipitation and further re-dissolving of dry residues (**Figure 3**).

How to deal with 'unknown' molecules

Untargeted metabolomics approaches also provide quantification of metabolites whose chemical identity has not yet been identified, though can be reproducibly detected and quantified. These metabolites, the so-called *unknowns*, have a big impact on biomedical research as many of them show strong correlations with clinical phenotypes[19-21] but their use as functional biomarkers is clearly very limited. The majority of studies ignores the unknowns either by platform design[22] or at the analysis stage[23,24]. Others have tried to solve the problem by employing different computer based methods to try and uncover the unknown true identity. Genetic data is particularly useful in this sense as several unknown metabolites are associated with genetic variation[5,7] and the strong genetic association can provide some clue on the metabolic pathway the unknown is most likely to be in. Building on this, Krumsiek and colleagues, developed a novel functional metabolomics method to predict unknown metabolites identity, integrating high-throughput genotyping data, metabolomics data and pathways information derived from the literature[25]. The advantage of their method is that it can be applied to metabolomics database obtain from commercial service, for which retention times, isotope patterns and fragmentation are not readily available.

ANALYTICAL STRENGTHS AND LIMITATIONS

Statistical techniques: data reduction, pathway analysis

Different statistical techniques can be employed in the analysis of metabolomics data (**Figure 4**)[26].

Usually the starting point is quality control of the metabolite data which may involve several steps including (i) data normalisation to remove/reduce the unwanted overall variation in the spectral data, (ii) standardising the data so that the data distribution has a mean of 0 when data is measured in a relative manner and (iii) scaling to account for different levels of concentration across metabolites with a focus on those with medium/ low-concentration. After this univariate and multivariate statistical analyses are performed to identify candidate biomarkers. Univariate analyses, such as linear model, t-test or ANOVA, examine each variable separately and are particularly adequate when the aim is to identify metabolites associated with disease pathogenesis . Stringent multiple testing correction is normally applied (Bonferroni/False Discovery Rate), which can obscure some biologically relevant observation. However, despite this obvious limitations including type 2 errors, they produce very solid and reproducible results.

Multivariate analysis is powerful for classification and prediction since one biomarker alone might not be by itself specific for a given trait. Multivariate methods range from principal component analysis (PCA), partial least squares - discriminant analysis (PLS-DA), penalised regression, random forest and support vector machine (SVM)[27]. PCA is particularly useful to get an overview of the data and to define previously unknown relationship between metabolites in a given physiological state. On the other hand, supervised methods such as PLS SVM can be used to build a model for classification. Most of the multivariate methods need full data matrices which introduces the problem of imputation.

Independent replication and functional studies are especially important to validate the results. While biomarkers are useful for risk prediction and health assessment, the value of a biomarker increases substantially if it is causally involved in the manifestation of a condition such as OA/RA etc. Causality can only be proven experimentally but multi-omics data allow for interesting modelling and statistical inference of causality. The genetic effect is by definition causal so combining genetic and metabolomics data may allow to infer causality networks of the phenotypes. Metabolic associations are difficult to understand in complex traits[28]. Using omics technologies helps to

decipher the biological meaning of otherwise obscure statistical associations. Benefits of the use of pathway and ontological analyses of genomics data have been extensively presented[29,30]. Analysis of biological pathways using the Database for Annotation, Visualisation and Integrated Discovery (DAVID) v6.7[31] and the Ingenuity Pathway analysis (IPA)[32] can be used to identify potential pathways determining metabolite signatures associated with particular conditions. Indeed these functional annotation tools used for genomic/genetic studies can be applied to the genes identified via GWAS or transcriptomics as associated with the metabolites under study and hence provide information on the pathways linked to these disease or drug response associated metabolites.

Open-source data repositories

Several open-source data repositories, as well as open access databases containing chemical, biological, molecular information about metabolites are now available. These include the Human Metabolon Database (HMDB)[10], PubChem[33], the Kyoto Encyclopedia of Genes and Genomes (KEGG)[34], the Metabolic Pathway Database (MetaCyc)[35], the Chemical Entities of Biological Interest database (ChEBI)[36], MetaboLIGHTS[37] and the metabolomics genome-wide association studies (GWAS http://mips.helmholtz-muenchen.de/proj/GWAS/gwas/) server, among others.

Moreover, web services have been developed for metabolomic data processing, analysis and annotation (e.g. MetaboAnalyst http://www.metaboanalyst.ca/faces/home.xhtml)

METABOLOMICS AND GENOMICS FOR RHEUMATOLOGIC TRAITS

A number of studies in the past few years have mapped genetic variation to plasma, serum and urine metabolites and have highlighted the genetic influence on these traits[4-9]. Also, using metabolomics as a readout of molecular phenotypes allows the discovery of formerly undetected associations between diseases, genes and metabolic pathways. The number of metabolomics studies has

exponentially increased (**Figure 5**) with over 4000 papers published in 2015, a quota not much lower than that of GWAS papers.

Metabolome

In the past few years various studies investigated the associations of rheumatic diseases with human metabolites (**Table 1** and **Supplementary Table 1**). These studies have shown that patients can be clearly distinguished from healthy controls based on the metabolic profile of blood[22] as well as urine[38] . Using a panel of 52 metabolites measured in circulating blood Madsen and colleagues[39] could accurately distinguish between RA patients and healthy controls as well as PsA patients. The authors found that patients could be distinguished from controls by decreased levels of (branched-chain) amino acids[20,40], increased levels of 3-hydroxybutyrate and lactate[41] and significant changes of energy metabolism[42,43]. Interestingly, similar metabolic changes also correlated with disease severity and response to treatment[44,45]. For instance, Kapoor and collaborators[3] predicted the response to anti-TNF therapy based on urine metabolite profiles with a sensitivity of 88.9% and specificity of 85.7%. Metabolomics studies also helped to identify causal mechanisms of disease, e.g. by establishing the role of pro-inflammatory estrogens[46], as well as the estrogens receptor alpha[47], which is activated by the metabolite 27-hydrocholesterol, in arthritis. Jiang and colleagues have performed metabolomic profiling in order to identify the global metabolic defects associated with four different types of arthritis including OA, RA, gout and ankylosing spondylitis, compared to healthy controls. They found a common metabolic defect that is due to joint inflammation and lesions, as well as a unique metabolic signature for each type of arthritis. This highlights the potential of metabolomics for biomarker discovery and treatment stratifications. However, many of the studies based their conclusions on small sample sizes, lacked replication and stringent correction for multiple testing. Future studies should address these issues to find reliable biomarkers of disease progression and response to treatment.

Genetics

There are three main aims behind any genetics study. The first is to identify proteins and pathways that are crucial to disease pathophysiology (e.g. OA) and increase our understanding of the condition. The second is to identify clinically relevant drug targets that can be manipulated pharmaceutically to prevent the disease (e.g. osteoporotic fractures). Finally, the third aim is to identify a set of variants that can be measured to determine which subject is at high risk (e.g. of future fracture). GWASes look for the association between common genetic variants -single nucleotide polymorphisms (SNPs), copy number variations (CNVs), indels - and specific phenotypes. In the past few years large-scale human GWASes of OA (>18000 samples[48]), RA (> 100000 samples[49]) have identified a substantial number of loci with several common pathway associated with prevalence, severity and progression of rheumatologic disease. More than 50 GWA studies found a total of 303 SNPs in 186 genes across all chromosomes that are associated with at least one rheumatic disease (**Figure 6**). These studies revealed that rheumatic diseases share common genetic risk factors, such as mutations in the *tumor necrosis factor, alpha-induced protein 3* (*TNFAIP3*) gene, which are associated with RA[49] , PsA[50] and SLE[51] The common genetic basis of these diseases is a likely reason for their comorbidities[52]. Common genetic causes also lead to similar symptoms, thus hindering the precise diagnosis of these diseases. Metabolomics is closer to the actual disease phenotype (**Figure 1**) and can therefore be used to differentiate between diseases where genetics cannot. However, in many cases the molecular links between the genes/biochemical pathway and the disease remain unknown. Animal models have been performed to explore the mechanism underlying some of the genes identified for rheumatologic diseases. Studies in mice have been able to show the effect of *TNFAIP3* on the inflammosome[53], while collagen-induced arthritis (CIA) mice were used to investigate the role of *IGFBP3* in NF-κB signalling and its subsequent effect on RA[54]. Work on *RUNX2* knockout mice explored their effect on OA by causing chondrocyte hypertrophy[55]. Researching the mechanisms of arthritis-related genes facilitates the discovery of drug targets, such as *RUNX2*[56] and *IGFBP3*[56]. However, the lack of functional annotation for many genes adds another

limitation to genetics studies. Metabolomics can help overcoming this shortcoming by inferring genes functions[57] and thus can provide further insights in potentially involved pathways.

Some of the loci identified via GWAS meta-analysis[48] lie in pathways known to be important in cartilage and bone physiology such as *CHST11*, *RUNX2* and *EGFBP3*, while in other cases the gene function is unknown. In cases whenthe aetiology of the link between a metabolite and disease onset or progression is known, it becomes feasible to make use of the metabolite in clinical research. On the other hand, this is not always needed for diagnosis and prognosis, as the is the case with anti-citrullinated protein antibodies (ACPA) as a biomarker for RA, which are used in clinical routine in spite of a lack of their role in disease pathogenesis[58] .

GWASes for the metabolome

A number of studies in the past few years have mapped genetic variation to plasma, serum and urine metabolites and have highlighted the genetic influence on these traits[4-8].

Using metabolomics as a readout of molecular phenotypes allows the discovery of formerly undetected associations between diseases, genes and metabolic pathways. While biomarkers are useful for risk prediction and health assessment, the value of a biomarker increases substantially if it is causally involved in the manifestation of any arthritis condition. Combining genetic and metabolomics data allows for interesting modelling and statistical inference of causality.

For instance, Mendelian randomisation is a statistical technique that allows to use genetic variation related to the risk factor of interest to re-assess observational estimates. The genetic variant acts as a proxy for the risk factor and the random allocation of genes during gamete production and fertilisation (Mendel Second law) is used as a natural experiment to show causation[59]. Being able to establish a causal link between a metabolite and disease or disease progression means that altering metabolite levels is directly relevant to disease management, and hence a target for intervention, rather than a simple by-product of disease. This is particularly relevant in the case of diseases such

as OA where no disease modifying drugs exist to date in spite of the very high prevalence and unmet medical burden.

GWA studies have shown that, with the exception of HLA associations with autoimmune disorders, effect sizes are modest and large sample sizes, usually in the range of thousands are needed. On the other hand, current metabolomic studies for rheumatological diseases have been carried out at most a few hundred individuals (see **Table 1** and **Supplementary Table 1**).

Utilizing both genetic knowledge and metabolomics is achievable. By querying GWAS data for variants that regulate levels of metabolites and by developing databases of metabolomics profiling for various conditions on genotyped individuals it will be possible to establish the links between genetic susceptibility and changes in metabolomics profiles in serum, urine, joint and bone tissue. To do this it will require being able to replicate findings in sufficiently large clinical cohorts of rheumatologically relevant traits assayed for the same metabolite panels and measured at comparable time points (e.g. pre/ post treatment). As these data become available it will become possible to point out the genetic basis of metabolite links to disease progression or response to therapy and hence to understand the molecular pathways underlying these processes.

IMPLICATIONS FOR THE FUTURE

In addition, metabolomic analyses can provide information on current medication use (i.e. metabolites derived from specific drugs) and levels of these metabolites, if dosage is known, can inform variation in drug response between patients or help track compliance in clinical trials. They can also inform on smoking, as cotinine, a metabolic product of nicotine, is commonly measured by metabolomic panels.

Metabolomics concentrates on small molecules. Other dynamic markers can also be integrated with this kind of data and should enrich our understanding of disease pathogenesis and improve disease diagnosis, prognosis and management in the future. We discuss two of these technologies: proteomics and metagenomics.

PROTEOMICS: The proteome includes all expressed proteins in a cell, tissue or organism and as such is a dynamic reflection of both genes and the environment. The human proteome map currently includes more than 84% of the annotated protein-coding genes in humans and 30,057 proteins derived from 17 adult and 6 foetal tissues (http://www.humanproteomemap.org/). It is a promising field for biomarker search as proteins are most likely to be ubiquitously affected in disease and disease response. Protein activity is indeed modulated by many factors besides the expression level of the relevant gene. Some direct and indirect approaches show promising results. For instance, emerging work on IgG glycosylation (the only post-translational modification that produces structural changes to proteins) has shown that patients with lupus have a significantly altered IgG glycome with decrease immunosuppressive action of circulating immunoglobulins compared to controls.[60]

METAGENOMICS: Despite GWASes and GWAS meta-analyses of increasingly large samples, the variants identified for RA, for example, account for less than a third of the estimated heritability of RA and less than 5% of OA. This "missing heritability" may be explained in a number of ways including structural variation, rare variants, or by environmental factors influenced by host genetics. The microbial organisms resident on and within a human host – the microbiome - is one such potential contributor[61]. Microbes produce a range of enzymes, chemicals, hormones and vitamins that can potentially interact with our bodies. They also produce up to 1/3 of circulating metabolites and their presence in the gut is partly influenced by human genetic factors[61]. Under physiological conditions, there is a balance between the intestinal bacteria and the host. Disruption of this intricate system (dysbiosis) has been implicated in many human diseases including RA[62]. Because the microbiome may be manipulated relatively easily with dietary or probiotic interventions it constitutes an attractive target for therapeutic intervention[63].

CONCLUSIONS

To date no published studies exist that combine a genetic and metabolomics approach to understand the pathogenesis of rheumatologic diseases. To a large extent this reflects the sparsity of metabolomics studies for the various forms of arthritis and the challenges posed by collecting relevant tissues (e.g. joint, bone, synovium) for cases and controls. Whereas genetics has delivered a remarkable wealth of knowledge on some types of arthritis, it has serious drawbacks, not least its non dynamic nature, and the modest effect sizes of any genes discovered (outside of the HLA for auto-immune disorders) and thus requires very large sample sizes.

Harnessing the power of both, genetic knowledge and metabolomics, is however doable. By querying GWAS data for variants that regulate levels of metabolites and by developing databases of metabolomics profiling for various conditions on genotyped individuals it will be possible to establish the links between genetic susceptibility and changes in metabolomics profiles in serum, urine, joint and bone tissue. This could help target pathways most likely to change over time or to improve early diagnosis or risk of progression. As the price of metabolomics decreases and it becomes increasingly feasible to measure hundreds of metabolites accurately, it will be used increasingly in rheumatology to assess drug levels and efficacy and could rapidly replace the more expensive clinical tests used today.

KEY POINTS

- Large scale genetic association analyses have identified variants underlying genetic architecture of various forms of arthritis but, with the exception of HLA loci, effect sizes are small.
- Metabolomics is a promising field both to investigate the molecular pathogenesis of rheumatologic diseases and to track response to therapy. Few modest sized studies have been conducted to date.
- Metabolomics can accurately measure drug metabolites like ibuprofen and paracetamol and other pain medications.
- In order for metabolomics to yield its full potential the genetic contribution to metabolites and to disease processes needs to be combined.
- Combining different omics including microbiome and proteomics may aid to increase the knowledge of pathways and diseases

REFERENCES

1 **Nicholson, J. K. & Lindon, J. C. Systems biology: Metabonomics. *Nature* 455, 1054-1056, doi:10.1038/4551054a (2008).**

2 **Tektonidou, M. G. & Ward, M. M. Validity of clinical associations of biomarkers in translational research studies: the case of systemic autoimmune diseases. *Arthritis research & therapy* 12, R179, doi:10.1186/ar3143 (2010).**

3 **Kapoor, S. R. *et al.* Metabolic profiling predicts response to anti-tumor necrosis factor alpha therapy in patients with rheumatoid arthritis. *Arthritis and rheumatism* 65, 1448-1456, doi:10.1002/art.37921 (2013).**

4 **Suhre, K. *et al.* A genome-wide association study of metabolic traits in human urine. *Nature genetics* 43, 565-569, doi:10.1038/ng.837 (2011).**

5 **Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 477, 54-60, doi:10.1038/nature10354 (2011).**

6 **Illig, T. *et al.* A genome-wide perspective of genetic variation in human metabolism. *Nature genetics* 42, 137-141, doi:10.1038/ng.507 (2010).**

7 **Shin, S. Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nature genetics* 46, 543-550, doi:10.1038/ng.2982 (2014).**

8 **Rhee, E. P. *et al.* A genome-wide association study of the human metabolome in a community-based cohort. *Cell metabolism* 18, 130-143, doi:10.1016/j.cmet.2013.06.013 (2013).**

9 **Kettunen, J. *et al.* Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature genetics* 44, 269-276, doi:10.1038/ng.1073 (2012).**

10 **Wishart, D. S. *et al.* HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic acids research* 41, D801-807, doi:10.1093/nar/gks1065 (2013).**

11 **Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature reviews. Molecular cell biology* 13, 263-269, doi:10.1038/nrm3314 (2012).**

12 **Lewis, G. D., Asnani, A. & Gerszten, R. E. Application of metabolomics to cardiovascular biomarker and pathway discovery. *Journal of the American College of Cardiology* 52, 117-123, doi:10.1016/j.jacc.2008.03.043 (2008).**

13 **Malet-Martino, M. & Holzgrabe, U. NMR techniques in biomedical and pharmaceutical analysis. *Journal of pharmaceutical and biomedical analysis* 55, 1-15, doi:10.1016/j.jpba.2010.12.023 (2011).**

14 **Griffiths, W. J. *et al.* Targeted metabolomics for biomarker discovery. *Angew Chem Int Ed Engl* 49, 5426-5445, doi:10.1002/anie.200905579 (2010).**

15 **Koal, T. & Deigner, H. P. Challenges in mass spectrometry based targeted metabolomics. *Current molecular medicine* 10, 216-226 (2010).**

16 **Guma, M., Tiziani, S. & Firestein, G. S. Metabolomics in rheumatic diseases: desperately seeking biomarkers. *Nature reviews. Rheumatology* 12, 269-281, doi:10.1038/nrrheum.2016.1 (2016).**

17 **Feng, X., Liu, X., Luo, Q. & Liu, B. F. Mass spectrometry in systems biology: an overview. *Mass spectrometry reviews* 27, 635-660, doi:10.1002/mas.20182 (2008).**

18 **Kind, T. *et al.* FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Analytical chemistry* 81, 10038-10048, doi:10.1021/ac9019522 (2009).**

19 **Armstrong, A. W. *et al.* Metabolomics in psoriatic disease: pilot study reveals metabolite differences in psoriasis and psoriatic arthritis. *F1000Research* 3, 248, doi:10.12688/f1000research.4709.1 (2014).**

20 **Zabek, A. *et al.* Application of (1)H NMR-based serum metabolomic studies for monitoring female patients with rheumatoid arthritis. *Journal of pharmaceutical and biomedical analysis* 117, 544-550, doi:10.1016/j.jpba.2015.10.007 (2016).**

21 **Loeser, R. F. *et al.* Association of urinary metabolites with radiographic progression of knee osteoarthritis in overweight and obese adults: an exploratory study. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society*, doi:10.1016/j.joca.2016.03.011 (2016).**

22 **Zhang, W. *et al.* Metabolomic analysis of human plasma reveals that arginine is depleted in knee osteoarthritis patients. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society* 24, 827-834, doi:10.1016/j.joca.2015.12.004 (2016).**

23 **Menni, C. *et al.* Metabolomic identification of a novel pathway of blood pressure regulation involving hexadecanedioate. *Hypertension* 66 (2015).**

24 **Menni, C. *et al.* Circulating levels of antioxidant vitamins correlate with better lung function and reduced exposure to ambient pollution. *American journal of respiratory and critical care medicine* 191, 1203-1207, doi:10.1164/rccm.201411-2059LE (2015).**

25 **Krumsiek, J. *et al.* Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. *PLoS genetics* 8, e1003005, doi:10.1371/journal.pgen.1003005 (2012).**

26 **Xi, B., Gu, H., Baniasadi, H. & Raftery, D. Statistical analysis and modeling of mass spectrometry-based metabolomics data. *Methods Mol Biol* 1198, 333-353, doi:10.1007/978-1-4939-1258-2_22 (2014).**

27 **Madsen, R., Lundstedt, T. & Trygg, J. Chemometrics in metabolomics--a review in human disease diagnosis. *Analytica chimica acta* 659, 23-33, doi:10.1016/j.aca.2009.11.042 (2010).**

28 **Nikolic, S. B., Sharman, J. E., Adams, M. J. & Edwards, L. M. Metabolomics in hypertension. *Journal of hypertension* 32, 1159-1169, doi:10.1097/HJH.0000000000000168 (2014).**

29 **Manoli, T. *et al.* Group testing for pathway analysis improves comparability of different microarray datasets. *Bioinformatics* 22, 2500-2506, doi:10.1093/bioinformatics/btl424 (2006).**

30 **Draghici, S., Khatri, P., Martins, R. P., Ostermeier, G. C. & Krawetz, S. A. Global functional profiling of gene expression. *Genomics* 81, 98-104 (2003).**

31 **Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4, 44-57, doi:10.1038/nprot.2008.211 (2009).**

32 **Kramer, A., Green, J., Pollard, J., Jr. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 30, 523-530, doi:10.1093/bioinformatics/btt703 (2014).**

33 **Wheeler, D. L. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic acids research* 36, D13-21, doi:10.1093/nar/gkm1000 (2008).**

34 **Kanehisa, M. *et al.* From genomics to chemical genomics: new developments in KEGG. *Nucleic acids research* 34, D354-357, doi:10.1093/nar/gkj102 (2006).**

35 **Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research* 44, D471-480, doi:10.1093/nar/gkv1164 (2016).**

36 **Hastings, J. *et al.* ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic acids research* 44, D1214-1219, doi:10.1093/nar/gkv1031 (2016).**

37 **Kale, N. S. *et al.* MetaboLights: An Open-Access Database Repository for Metabolomics Data. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* 53, 14 13 11-18, doi:10.1002/0471250953.bi1413s53 (2016).**

38 **Yan, B. *et al.* Urinary metabolomic study of systemic lupus erythematosus based on gas chromatography/mass spectrometry. *Biomedical chromatography : BMC*, doi:10.1002/bmc.3734 (2016).**

39 **Madsen, R. K. *et al.* Diagnostic properties of metabolic perturbations in rheumatoid arthritis. *Arthritis research & therapy* 13, R19, doi:10.1186/ar3243 (2011).**

40 **Zhai, G. *et al.* Serum branched-chain amino acid to histidine ratio: a novel metabolomic biomarker of knee osteoarthritis. *Annals of the rheumatic diseases* 69, 1227-1231, doi:10.1136/ard.2009.120857 (2010).**

41 **Young, S. P. *et al.* The impact of inflammation on metabolomic profiles in patients with arthritis. *Arthritis and rheumatism* 65, 2015-2023, doi:10.1002/art.38021 (2013).**

42 **Wu, T. *et al.* Metabolic disturbances associated with systemic lupus erythematosus. *PloS one* 7, e37210, doi:10.1371/journal.pone.0037210 (2012).**

43 **Perl, A. *et al.* Comprehensive metabolome analyses reveal -acetylcysteine-responsive accumulation of kynurenine in systemic lupus erythematosus: implications for activation of the mechanistic target of rapamycin. *Metabolomics* 11, 1157-1174, doi:10.1007/s11306-015-0772-0 (2015).**

44 **Kamleh, M. A. *et al.* LC-MS metabolomics of psoriasis patients reveals disease severity-dependent increases in circulating amino acids that are ameliorated by anti-TNFalpha treatment. *Journal of proteome research* 14, 557-566, doi:10.1021/pr500782g (2015).**

45 **Priori, R. *et al.* (1)H-NMR-Based Metabolomic Study for Identifying Serum Profiles Associated with the Response to Etanercept in Patients with Rheumatoid Arthritis. *PloS one* 10, e0138537, doi:10.1371/journal.pone.0138537 (2015).**

46 **Schmidt, M. *et al.* Estrone/17beta-estradiol conversion to, and tumor necrosis factor inhibition by, estrogen metabolites in synovial cells of patients with rheumatoid arthritis and patients with osteoarthritis. *Arthritis and rheumatism* 60, 2913-2922, doi:10.1002/art.24859 (2009).**

47 **Umetani, M. *et al.* The cholesterol metabolite 27-hydroxycholesterol promotes atherosclerosis via proinflammatory processes mediated by estrogen receptor alpha. *Cell metabolism* 20, 172-182, doi:10.1016/j.cmet.2014.05.013 (2014).**

48 **Zeggini, E. *et al.* Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* 380, 815-823, doi:10.1016/S0140-6736(12)60681-3 (2012).**

49 **Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506, 376-381, doi:10.1038/nature12873 (2014).**

50 **Yin, X. *et al.* Genome-wide meta-analysis identifies multiple novel associations and ethnic heterogeneity of psoriasis susceptibility. *Nature communications* 6, 6916, doi:10.1038/ncomms7916 (2015).**

51 **Yang, W. *et al.* Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. *American journal of human genetics* 92, 41-51, doi:10.1016/j.ajhg.2012.11.018 (2013).**

52 **Michaud, K. & Wolfe, F. Comorbidities in rheumatoid arthritis. *Best practice & research. Clinical rheumatology* 21, 885-906, doi:10.1016/j.berh.2007.06.002 (2007).**

53 **Vande Walle, L. *et al.* Negative regulation of the NLRP3 inflammasome by A20 protects against arthritis. *Nature* 512, 69-73, doi:10.1038/nature13322 (2014).**

54 **Lee, H. S. *et al.* Regulation of apoptosis and inflammatory responses by insulin-like growth factor binding protein 3 in fibroblast-like synoviocytes and experimental animal models of rheumatoid arthritis. *Arthritis Rheumatol* 66, 863-873, doi:10.1002/art.38303 (2014).**

55 **Kamekura, S. *et al.* Contribution of runt-related transcription factor 2 to the pathogenesis of osteoarthritis in mice after induction of knee joint instability. *Arthritis and rheumatism* 54, 2462-2470, doi:10.1002/art.22041 (2006).**

56 **Schminke, B., Trautmann, S., Mai, B., Miosge, N. & Blaschke, S. Interleukin 17 inhibits progenitor cells in rheumatoid arthritis cartilage. *European journal of immunology* 46, 440-445, doi:10.1002/eji.201545910 (2016).**

57 **Adamski, J. Genome-wide association studies with metabolomics. *Genome medicine* 4, 34, doi:10.1186/gm333 (2012).**

58 **Conigliaro, P. *et al.* Autoantibodies in inflammatory arthritis. *Autoimmunity reviews* 15, 673-683, doi:10.1016/j.autrev.2016.03.003 (2016).**

59 **Smith, G. D. & Ebrahim, S. Mendelian randomization: prospects, potentials, and limitations. *International journal of epidemiology* 33, 30-42, doi:10.1093/ije/dyh132 (2004).**

60 **Vuckovic, F. *et al.* Association of systemic lupus erythematosus with decreased immunosuppressive potential of the IgG glycome. *Arthritis Rheumatol* 67, 2978-2989, doi:10.1002/art.39273 (2015).**

61 **Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* 159, 789-799, doi:10.1016/j.cell.2014.09.053 (2014).**

62 **Lin, P. *et al.* HLA-B27 and human beta2-microglobulin affect the gut microbiota of transgenic rats. *PloS one* 9, e105684, doi:10.1371/journal.pone.0105684 (2014).**

63 **Vellucci, R. Heterogeneity of chronic pain. *Clinical drug investigation* 32 Suppl 1, 3-10, doi:10.2165/11630030-000000000-00000 (2012).**

64 **Baurecht, H. *et al.* Genome-wide comparative analysis of atopic dermatitis and psoriasis gives insight into opposing genetic mechanisms. *American journal of human genetics* 96, 104-120, doi:10.1016/j.ajhg.2014.12.004 (2015).**

65 **Li, C. *et al.* Genome-wide association analysis identifies three new risk loci for gout arthritis in Han Chinese. *Nature communications* 6, 7041, doi:10.1038/ncomms8041 (2015).**

66 **Hom, G. *et al.* Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *The New England journal of medicine* 358, 900-909, doi:10.1056/NEJMoa0707865 (2008).**

67 **Lamers, R. J. *et al.* Identification of an urinary metabolite profile associated with osteoarthritis. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society* 13, 762-768, doi:10.1016/j.joca.2005.04.005 (2005).**

68 **Mickiewicz, B. *et al.* Metabolic analysis of knee synovial fluid as a potential diagnostic approach for osteoarthritis. *Journal of orthopaedic research : official publication of the Orthopaedic Research Society* 33, 1631-1638, doi:10.1002/jor.22949 (2015).**

69 **Zhou, J. *et al.* Exploration of the serum metabolite signature in patients with rheumatoid arthritis using gas chromatography-mass spectrometry. *Journal of pharmaceutical and biomedical analysis*, doi:10.1016/j.jpba.2016.02.004 (2016).**

70 **Ouyang, X., Dai, Y., Wen, J. L. & Wang, L. X. (1)H NMR-based metabolomic study of metabolic profiling for systemic lupus erythematosus. *Lupus* 20, 1411-1420, doi:10.1177/0961203311418707 (2011).**

71 **Yan, B. *et al.* Serum metabolomic profiling in patients with systemic lupus erythematosus by GC/MS. *Modern rheumatology / the Japan Rheumatism Association*, 1-9, doi:10.3109/14397595.2016.1158895 (2016).**

72 **Sitter, B., Johnsson, M. K., Halgunset, J. & Bathen, T. F. Metabolic changes in psoriatic skin under topical corticosteroid treatment. *BMC dermatology* 13, 8, doi:10.1186/1471-5945-13-8 (2013).**

73 **Liu, Y. *et al.* A metabolic profiling analysis of symptomatic gout in human serum and urine using high performance liquid chromatography-diode array detector technique. *Clinica chimica acta; international journal of clinical chemistry* 412, 2132-2140, doi:10.1016/j.cca.2011.07.031 (2011).**

74 **Albrecht, E. *et al.* Metabolite profiling reveals new insights into the regulation of serum urate in humans. *Metabolomics* 10, 141-151, doi:10.1007/s11306-013-0565-2 (2014).**

ACKNOWLEDGEMENTS

TwinsUK was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR) Clinical Research Facility at Guy's & St Thomas' NHS Foundation Trust and NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. Tim Spector is an NIHR senior Investigator and is holder of an ERC Advanced Principal Investigator award. This work is also supported by the MRC AimHY (MR/M016560/1) grant.

DISCLOSURE

The authors declare no conflict of interest

AUTHOR BIOGRAPHIES

Cristina Menni received her Master in mathematics from the University of Cambridge and her PhD in statistics from the University of Milan-Bicocca where she specialised in statistical genetics. As a? research fellow at the Department of Twin Research & Genetic Epidemiology at King's College London, her research is focused on identifying novel molecular markers associated with age related diseases using metabolomics in conjunction with several other "omics" datasets.

Jonas Zierer is a PhD student at King's College London, UK, in a joint project with the Helmholtz-Zentrum Munich, Germany. After graduating from his masters in Bioinformatics at the University of Munich he focused his research on metabolic signature that underlie age-related diseases.

Ana M Valdes received her M.A. and Ph.D. from the University of California at Berkeley where she specialized in Genetic Epidemiology. A Reader and Associated Professor at the University of

Nottingham, her work focuses on deciphering the molecular mechanisms using metabolomics and genetics related to age-related phenotypes and to diseases including osteoarthritis and gout.

Tim D Spector M.D. Ph.D. is a consultant Rheumatologist and Professor in Genetic Epidemiology head of the Department of Twins Research at King's College London, he is the founder of the TwinsUK cohort, one of the most extensively phenotyped cohorts in the world regarding omic measurements. He has published extensively in the fields of genetics, metabolomics and metagenomics.

Figure legends.

Figure 1. Metabolomic profiling as a tool in the study of rheumatological pathologies. The metabolites in circulation or found in tissues reflect on the one hand the genetic background and the gene expression in an individual, but also their lifestyle, diet, medication and exposures and the pathogenic processes ongoing. An integration with genetic profiling can help differentiate between molecular changes seen in a given joint disease derived from modifiable factors (i.e. dynamic changes that may respond to diet or medication) and those that are directly linked to an individual's genetic profile. They may also be used to track progression of disease and of treatment

Figure 2. Two main approaches exist for carrying out metabolomic profiling: targeted and untargeted. The scope of these two types of analysis is different and they both have advantages and disadvantages. Untargeted metabolomics profiling is almost exclusively carried out using MS methods

Figure 3. Metabolomic profiling can be carried out in almost any type of biospecimen. The availability of joint or bone tissues is a major challenge. On the other hand, data from serum or urine may not be directly relevant to specific pathogenic processes.

Figure 4. Statistical approaches to analyse metabolomics data

Figure 5. Publication statistics. Number of publications that were indexed in pubmed per year since 2000 referring to GWASs (keywords “GWAS”, “genome-wide association”) and metabolomics (keywords “metabolomics”, “metabonomics”, “metabolic profiling”), respectively.

Figure 6. Genetic basis of rheumatoid diseases

The figure summarises genetic associations with rheumatoid diseases listed in the GWAS catalogue. We included all SNPS that are associated with one of the diseases of interest at genome-wide significant (p<5*10-8), leaving 303 SNPs from 56 studies. SNP names and the associated genes were added for the SNPs with the highest p-values in each bin of 10^7bp. The associations were retrieved

from the GWAS catalogue (https://www.ebi.ac.uk/gwas) using the catalogue traits “osteoarthritis” and “osteoarthritis, knee” for OA, “psoriatic arthritis” and “psoriasis” for PsA, “rheumatoid arthritis” for RA, “gout” and “gout, urate measurement” for gout and “systemic lupus erythematosus”, “interferon alpha measurement, systemic lupus erythematosus”, “neonatal systemic lupus erythematosus", “systemic lupus erythematosus, lupus nephritis” and “systemic scleroderma, systemic lupus erythematosus” for lupus. The retrieved studies include (but are not limited to) studies on OA using 7410 cases and 11009 controls[48], on PsA based on 4212 cases and 8032 controls[64], on RA based on 29880 cases and 73758 controls from different ancestries[49], on gout using 4275 cases and 6272 controls[65] and studies on SLE based on 1311 cases and 2240 controls from European ancestry[66] (and 1656 cases with 3394 controls with Chinese ancestry [51], respectively.

Table 1 – Examples of metabolomics studies investigating rheumatic diseases. Further details on the key studies are found in Supplementary Table 1.

Disease	# studies	Study design	# metabolites studied	Platform used	Fluid/tissue studied	Key studies
OA	61	Case control (range n=50 to 300 per group), progressors vs non progressors (n=22 per group)	Untargeted NMR signals* (>40,000) Targeted 55-228	Targeted NMR, untargeted NMR, GC-MS	Urine, plasma, serum knee synovial fluid	Lamers 2005 [67] Mickiewicz 2015 [68] Zhang 2016 [22] Loeser 2016 [21] Zhai 2010 [40]
RA	87	Response to anti-TNF treatment (n<=30), prospective study of early RA (n=216), case-control (n~55 per group)	Untargeted NMR signals* (>40,000) Targeted 194	Targeted NMR, untargeted NMR, GC-MS	serum	Priori 2015 [45] Zabek 2016 [20] Young 2013 [41] Zhou 2016 [69]
SLE	32	Case control (range n=20-80 per group)	Targeted 81-319	Targeted LC-MS Targeted GC-MS Targeted NMR	Serum, urine, peripheral blood lymphocytes	Young 2013 [41] Wu 2012 [42] Ouyang 2011 [70] Yan 2016 [38] Yan 2016 [71]
PsA	18	Case control (range n=10-32 per group) Response to therapy (range n=10-32) (anti TNF, steroids)	Targeted 9-354	Targeted GC-MS Targeted LC-MS Targeted NRM	Serum, urine, plasma, skin	Armstrong 2014 [19], Kapoor 2013 [3], Kamleh 2015 [44], Sitter 2013 [72]
gout	17	Case control (n=21) Population based urate levels	Targeted 27-355	Targeted LC-MS, GC-MS HPLC diode array detector	serum and urine	Liu 2011 [73], Albrecht 2014 [74]

***** number of data points used in the NMR data processing

Supplementary Table 1 – Examples of metabolomics studies investigating rheumatic diseases.

Pheno	# studies	Study design	# metabolites studied	Platform used	Fluid studied	Key Findings	REF
OA	61	45 cases and 47 controls	65000 NMR signals	NMR	Urine	- urine metabolite profiles discriminate between cases and controls - metabolite profiles differ by OA severity - changes in histidine metabolism	Lamers 2005 [67]
		55 cases, 13 (post-mortem) controls	55	NMR + GC-MS	knee synovial fluid	- fructose and citrate increased in cases, -acetylcarnitine, N-phenylacetylglycine, methionine, ethanol, creatine, malate, ethanolamine, 3-hydroxybutyrate and hexanoylcarnitine decreased in cases - multivariate analysis perfectly distinguishes between cases and controls	Mickiewicz 2015[68]
		64 cases and 45 controls, replicated in 72 cases and 76 controls	186	MS (Biocrates)	plasma	- 1 amino acid, 4 phosphatidylcholines and one sphingomyleine significantly different between cases and controls after replication - arginine levels of <=57mM predict OA cases with 98.3% sensitivity and 89% specificity	Zhang 2016[22]
		prospective study with 22 progressors and 22 age- and sex-matched non-progressors	228	NMR	urine	- urine metabolite levls distinguishbetween progressors and non-progressors prospectively at baseline as well as at follow-up (18 month later) - glycolate and hippurate are increased and trigonelline decreased in progressors - histidine is increased in progressors at follow up	Loeser 2016[21]
		123 cases and 299 controls, replicated in 76 cases and 100 controls	163 (and their ratios)	MS(Biocrates)	serum	- ratios of valine/histidine and xleucine/histidine consistenly increased in OA patients - other BCAA suggestively associated with OA	Zhai 2010[40]
RA	87	prospective study with 27 female patients to assess response to Etanercept	16000 NMR peaks	NMR	serum	- metabolic profiling allows perfect separation of responders and non-responders - branched-chain amino acids, alanine, glutamine, tyrosine, and glucose levels increased in responders; 3-hydroxybutyrate depleated in responders	Priori 2015[45]
		30 healthy controls and 20 RA cases (followed up after treatment with etanercept or adalimumab)	47 599 NMR peaks	NMR	serum	- perfect separation of cases and controls based on (amongst others) branched chain amino acids, lactate, 3-hydroxyisobutyrate and histidine - cases after treatment form new cluster, clearly distinguishable from cases as well as healthy controls	Zabek 2016[20]
		16 patients with established RA, 14 healthy controls and 216 patients with early RA (followed-		NMR	serum	- lipid profiles in early arthritis change with level of inflammation (based on C-reactive protein) - 3-hydroxybutyrate and lactate elevated in RA patients - discrimination of RA progressors and non-progressors possible based on Acetylglycine, methylguanidine and lactate, but with low accuracy	Young 2013[41]

		up longitudinally)					
		55 RA cases, 32 controls	194	GC-MS	serum	- amino acids depleated in RA cases compared to controls - several metabolites involved in glycolysis, TCA and Urea Cycle disturbed in RA cases	Zhou 2016[69]
SLE	32	42 healthy controls and 36 SLE cases in a double-blind placebo-controlled trial with N-acetylcysteine (NAC) treatment	258	LC-MS	peripheral blood lymphocytes	- major differenctes between cases and healthy controls in pyrimidine, pyruvate and pentose phosphat pathways - metabolic profiles also associated with disease activity, measured by SLEDAI, BILAG and FAS - Kynurenine reduced by treatment with N-acetylcysteine (NAC) compared to placebo	Young 2013[41]
		20 cases and 9 controls, replicated in 38 cases using different assays	319	GC-MS + LC-MS	serum	- reduced energy biogenesis in SLE cases, indicated by changes in glycolysis, TCA cycle, beta oxidation and amino acid metabolism - increased ocidative stress in cases: peroxidation of fatty acids, gluthatione	Wu 2012[42]
		64 SLE patients, 30 RA patients and 35 healthy controls		NMR	serum	- lower concentration of amino acids, sugar, glycerides and citrate in SLE and RA cases compared to controls - SLE cases characterised by dyslipidemia (elevated LDL, depleated HDL)	Ouyang 2011[70]
		28 cases, 47 controls	81	GC-MS	urine	- cases and controls distinguishable with reasonable quality - concentration of branched chain amino acid in urine differs significantly between cases and controls	Yan 2016[38]
		80 cases, 57 controls	98	GC-MS	serum	- clear separation possible between cases and healthy controls as well as active and inactive SLE cases - significant differences in amino acid (particularly branched-chain), TCA and lipid metabolism - 2-hydroxyisobutyrate levels elevated in active but not in inactive SLE cases	Yan 2016[71]
PsA	18	10 psoriasis cases, 10 cases with psoriasis and psoriatic arthritis and 10 healthy controls	144 known + 354 unknown	GC-MS	serum	- psoriasis cases have higher levels of alpha ketoglutaric acid and lower levels of asparagine anf glutamine compared to controls - psoriatic arthritis cases have higher levels of glucuronic acid compared to controls - higher concentration of lignoceric acid in psoriatic arthritis than patients with psoriasis	Armstrong 2014[19]
		longitudinal study of 16 RA cases and 20 PsA cases before and during anti-TNF		NMR	urine	- urine metabolite profiles distinguish between baseline (RA and PsA) samples and after treatment samples as well as between different treatments	Kapoor 2013[3]

		therapy					
		32 healthy controls, 32 mild psoriasis cases, 32 severe psoriasis cases (latter before and after treatment wit Etanercept)	94	LC-MS	plasma	- mainly amino acid levels differ between cases and controls and correlate with severity of disease - treatment with Etanercept reverses many differences between cases and controls	Kamleh 2015[44]
		10 cases before and after treatment with emollient and the corticosteroid clobetasol propionate ointment	9	NMR	skin	- levels of myo-inositol and glucose elevated and levels of choline decreased in psoriatic skin compared to uninvolved skin, indicating high proliferation rates - steroid treatment leads to assimilation of treated psoriatic skin to uninvolved skin	Sitter 2013[72]
gout	17	21 gout cases and 21 age- and sex-matched controls	27 peaks in serum and 39 in urine	High performance liquid chromatography-diode array detector	serum and urine	- clear separation of cases and controls possible with both, serum and urine metabolite profiles - creatinine, uric acid and tryptophan increased in serum of cases - hippuric acid and guanine decreased in urine of gout patients	Liu 2011[73]
		cross-sectional with 1764 individuals investigating serum urate levels	355	LC-MS, GC-MS (Metabolon)	serum	- using gaussian graphical models reveals three modules involved in serum urate regulation: purine metabolites, amino acids and steroids	Albrecht 2014[74]

The table summarises metabolomics studies investigating different rheumatic diseases. Studies were retrieved from Pubmed using the keywords "(metabolomics OR metabonomics OR metabolic profiling)" and "osteoartrithis" for OA, "rheumatoid arthritis" for RA, "lupus" for SLE, "psoriatic arthritis OR psoriasis" for PsA and gout. An exemplary selection of studies for each disease is briefly described.