# The Contribution of Transcriptomics to Biomarker Development in Systemic Vasculitis and SLE.


## Authors

Flint, Shaun M[1*], McKinney, Eoin F[1], Lyons, Paul A[1], Smith, Kenneth GC[1]

[1] Cambridge Institute of Medical Research, and

Department of Medicine

The University of Cambridge

Box 139, Addenbrooke's Hospital

Hills Road, Cambridge CB2 0XY

United Kingdom

*Corresponding Author

Dr Shaun Flint

e-mail: sf424@cam.ac.uk

phone: 01223 762 639

fax: 01223 336 846


## Running Title:

Transcriptomic Biomarkers in Rheumatology


## Keywords

gene expression; biomarker; interferon; Systemic Lupus Erythematosus; vasculitis

## Abbreviations

Systemic Lupus Erythematosus (SLE)

ANCA-associated vasculitis (AAV)

Peripheral blood mononuclear cells (PBMC)

Ribonucleic acid (RNA)

RNA sequencing (RNA-Seq)

Complementary deoxyribonucleic acid (cDNA)

Messenger ribonucleic acid (mRNA)

Quantitative polymerase chain reaction (qPCR)

Type 1 interferon inducible (IFN-I)

Enzyme-linked immunosorbent assay (ELISA)

Anti-neutrophil cytoplasmic antibodies (ANCA)

## Abstract

A small but increasing number of gene expression based biomarkers are becoming available for routine clinical use, principally in oncology and transplantation. These underscore the potential of gene expression arrays and RNA sequencing for biomarker development, but this potential has not yet been fully realized and most candidates do not progress beyond the initial report. The first part of this review examines the process of gene expression-based biomarker development, highlighting how systematic biases and confounding can significantly skew study outcomes. Adequate validation in an independent cohort remains the single best means of protecting against these concerns.

The second part considers gene-expression based biomarkers in Systemic Lupus Erythematosus (SLE) and systemic vasculitis. The type 1 interferon inducible gene signature remains by far the most studied in autoimmune rheumatic disease. While initially presented as an objective, blood-based biomarker of active SLE, subsequent research has shown that it is not specific to SLE and that its association with disease activity is considerably more nuanced than first thought. Nonetheless, it is currently under evaluation in ongoing trials of anti-interferon therapy. Other candidate markers of note include a prognostic CD8+ T-cell gene signature validated in SLE and ANCA-associated vasculitis, and a disease activity biomarker for SLE derived from modules of tightly correlated genes.

## Introduction

Over the last decade, high-throughput genomic technology has matured to the extent that we can now study tissue RNA abundance across the genome. In oncology, five prognostic, multi-gene expression-based biomarkers are currently available for clinical use in patients with breast cancer, with at least one of these having additionally been shown to predict benefit from adjuvant chemotherapy [1]. In cardiac transplantation, some centers have implemented a gene expression biomarker in place of surveillance biopsies to screen for graft rejection [2]. Together, these examples highlight the wide potential of gene expression profiling as a discovery tool for biomarker development. Its major advantage lies in its ability to quantify sample messenger RNA (mRNA) for a large fraction of known genes, resulting in a hypothesis-free dataset that can then be studied for truly novel markers and pathways. Usual practice is to combine the expression of multiple genes into a single biomarker or 'signature', which means the range of potential biomarkers is vast and can be tailored to clinical requirements.

The promise of gene expression-based biomarkers is, however, yet to be fully realised. For every such biomarker in the late stages of clinical development, many more published candidates never progress beyond the initial report [3]. The first part of this review examines in general terms the process of discovery and subsequent translation of gene expression-based biomarkers because an understanding of this process and of the reasons that so many fail is central to identifying the biomarkers most likely to translate into useful clinical tests.

The second part applies this knowledge in a critical appraisal of gene expression-based biomarker research in ANCA-associated vasculitis (AAV), Systemic Lupus Erythematosus (SLE) and related vasculitides. These diseases all feature dysregulation of various immune cell subsets, ultimately causing pathology through autoantibody production and tissue inflammation. Many of these cells are readily sampled in the peripheral blood as they traffic between lymph node and tissues, providing a rationale that gene expression profiling of blood is as likely to yield relevant patterns as the more technically challenging approach of profiling inflamed tissues.

Of gene-expression based biomarkers in rheumatology, the most studied is the type 1 interferon inducible (IFN-I) gene signature that was first described for SLE over a decade ago [4,5]. We also consider in detail a CD8+ T-cell prognostic gene signature shared between flaring SLE and AAV patients, and an SLE disease activity biomarker based on 'modules' of tightly correlated genes in peripheral blood mononuclear cells (PBMC) [6,7]. These three examples are chosen for their potential to change clinical practice, and because they highlight very distinct approaches to gene expression-based biomarker discovery. A number of other early-stage biomarkers are also covered briefly, but in recent years the field has become too broad for an exhaustive review to be informative.

## 1. Gene expression-based biomarker discovery.

### Biomarkers are measurable surrogates for underlying biological processes.

As the term 'biomarker' has become more pervasive, precision in its definition has been eroded. By biomarker, we mean "an objectively measurable characteristic that reflects underlying biological processes." [8] Clinically useful biomarkers are those that have been shown to correlate with important clinical parameters such as diagnosis, disease phenotype, prognosis, response to therapy or disease

activity. There is a complex interplay between biomarker development and research into disease pathogenesis: for example, a role for type 1 interferon in SLE was known well before the description of an IFN-I signature of SLE, but the first IFN-I signature papers were based on a non-hypothesis driven analysis of gene expression data. Here, the term gene signature and gene expression-based biomarker are overlapping concepts. In general, gene expression-based biomarkers guide decision-making by using the expression of a shortlist of genes (or 'signature') to derive either a continuous 'score' or to classify samples into distinct groups.

## An idealized biomarker development pathway is similar to that for a new drug.

In an attempt to increase the proportion of proposed biomarkers that translate to the clinic, various organisations (e.g. the Early Detection Research Network) have begun to codify a stepwise process for biomarker development that mirrors the phased approach for novel therapeutics [9]. This process always begins with the initial discovery of a candidate marker (Figure 1). When proposed gene expression-based biomarkers fail to progress, it can often be traced back to flaws in this initial discovery step. Common problems that arise include choices made during patient and sample selection, clinical annotation, experimental design, technical bias, and data analysis.

## Choosing patients and deciding sample types.

Consider a biomarker study for renal transplant tolerance, in which the whole blood gene expression profiles of a group of operationally tolerant transplant recipients off immunosuppression were compared to those from a matched group of immunosuppressed transplant recipients and a second group of healthy volunteers [10]. The study found that a signature dominated by B-cell expressed genes was able to differentiate tolerant recipients from those still receiving immunosuppression and proposed that a subset of these genes might form a useful tool for screening renal transplant recipients for tolerance. However, the cohorts also differ by presence or absence of immunosuppression, and the alternative possibility that this drives the B-cell signature needs to be considered. Obviously, few alternative study designs are feasible in a transplant tolerance study, and the authors present some

additional evidence that their signature may be tolerance-related, but this illustrates how significant baseline differences between cohorts have the potential to confound a gene expression study.

A related issue to consider is that of generalizability: to what extent can a candidate biomarker be applied to a population different to that used in its discovery? Can a biomarker derived from a population with flaring disease then be applied to a treated population encompassing a range of disease activities? These questions are relevant because it is often expedient to undertake biomarker discovery using a homogenous subset of active patients, minimizing the chance of within-cohort variation obscuring important between-group differences and allowing sample sizes to be kept small. Ultimately, the only way to demonstrate generalizability is to study biomarker performance in the intended population.

The type of sample chosen has a significant impact on the practicality and cost of a biomarker study. Using whole blood, PBMC or other mixed cell populations in preference to purified cell subsets is convenient, cheap and speeds translation. Whole blood, in particular, is a sample of convenience: it can be collected directly into a tube containing RNA-stabilizing reagents (e.g. PAXgene), allowing the sample to be kept at room temperature for up to two days [11]. Any resulting expression-based biomarker can be transformed into a workable clinical diagnostic by designing and validating an appropriate multiplex quantitative PCR (qPCR) panel. The key drawback of this approach is that gene expression signals seen in mixed samples often reflect changes in sample composition, something perhaps more easily determined by flow cytometry or histology. Examples include the ‘B-cell’ signature discussed above or the ‘immature granulocyte’ signature of SLE discussed below, reflecting increased B-cell and granulocyte proportions, respectively [5,10]. Using mixed samples also tends to obscure changes in gene expression within single populations, with the consequence that potentially informative biomarkers are missed [12]. While the barriers for translating a biomarker based on separated cell populations into a clinical diagnostic are higher, novel technologies such as RNAscope (Advanced Cell Diagnostics) have the potential to facilitate this process by combining fluorescence-based quantitation of cellular mRNA transcripts with flow cytometry to identify subpopulations [13].

## Experimental design and technical bias.

Microarrays remain the most commonly used technology for gene expression-based biomarker studies, although this will change as sequencing based technologies (i.e. RNA sequencing [RNA-Seq]) become more accessible. Modern Affymetrix microarrays contain upwards of 750,000 25-base-pair oligonucleotides synthesised in bulk directly on the microarray surface. Matching complementary DNA (cDNA) derived from the sample mRNA bind to their complementary probe on the array. Bound cDNA is detected by fluorescence – where greater fluorescence implies greater expression – and then mapped back to a specific oligonucleotide and ultimately gene by position on the microarray [14].

As with other high throughput technology, microarrays are sensitive to the precise conditions under which the assay was performed. Samples run separately will incorporate subtle assay-derived variation or 'batch effects' [15]. Studies that do not account for this by using standard protocols and ensuring assay runs are balanced across cohorts of interest are at risk of introducing systematic bias due to these batch effects, as happened in one prominent study of gene expression across ethnic groups [16]. This is of particular relevance where stored samples are accessed for a study: consider a biomarker study that is to use stored RNA samples from SLE patients and prospectively recruited matched healthy volunteers as a comparison cohort. Unless the healthy volunteer samples are processed in an identical fashion to the banked samples and arrayed together, technical differences between the two cohorts could easily be misinterpreted as biological variation [17].

## The bioinformatics of biomarker discovery.

### *Class prediction.*

Whilst there are almost as many informatic methods as there are biomarker studies, these will generally fall under one of two strategies: class prediction, or class discovery [18]. For class prediction, the researcher attempts to assign samples to known, clinically relevant classes using only the gene expression data (e.g. can I define whether a patient with SLE will respond to rituximab based on their B-cell gene expression?). It is the most commonly used strategy, but sensitive to biases in patient and sample selection. To minimize this possibility, class prediction is best done in a prospective study design where samples are processed without prior knowledge of their clinical phenotype. Moreover,

class prediction assumes that samples within the classes being predicted are homogenous. If not, then variation in gene expression within a class may obscure inter-class differences – a particular risk in heterogeneous diseases such as SLE.

Increasingly sophisticated algorithms are used for class prediction, but most perform two tasks: feature reduction, in which a subset of the most informative genes are identified, and then the class prediction exercise itself. This can be as simple as taking the mean expression of the chosen genes and using a pre-determined threshold to classify samples, or as complex as training a support vector machine [19].

### *Class discovery.*

Class discovery makes no prior assumptions about the samples and proceeds without knowledge of their clinical phenotype. Instead, the researcher attempts to discover structure (e.g. clusters of samples) intrinsic to the expression data itself. If structure is identified, then the clinical phenotype of each cluster is examined for potential correlates (e.g. prognosis, disease activity). In class discovery, researchers should explicitly demonstrate that any clusters identified are unlikely to be an artifact of random variation given that most clustering algorithms will identify clusters in a randomly generated sample [20].

### Independent validation is essential.

Regardless of whether a class prediction or class discovery approach is implemented, two further concerns arise. First, the likelihood of chance associations or clusters in a dataset containing tens of thousands of data points per sample is high, and in a hypothesis-free analysis any one of these spurious associations might be taken forward as a positive result. Second, there is a significant risk of overfitting, where multi-gene signatures are matched so closely to the data that they model technical and random variation instead of, or alongside, true biologic differences.

While careful experimental design is the best, if not the only, way to defend against bias and while there are increasingly sophisticated statistical techniques to account for multiple-testing and minimise over-fitting, most authors agree that the single most important protection against these types of error is prospective validation of candidate biomarkers in a properly independent population [21]. This will

inevitably occur downstream of the discovery process, either to validate a clinical assay developed from the candidate biomarker or to validate a biomarker's role in clinical decision-making. However, given the time and resources expended before these points are reached, it is increasingly expected that researchers build a validation component into the discovery process. The early validation of biomarkers is not done well, if at all [3]. The most common errors are inappropriate use of cross-validation and inadequately powered validation cohorts resulting in error rates with (usually unreported) confidence intervals that are so wide as to be meaningless [22]. Cross-validation is a statistical technique for undertaking biomarker validation using the original dataset, but unless implemented in a rigorous and unbiased fashion it is very prone to overestimating biomarker performance. As such, studies implementing cross-validation alone should be viewed with caution [23].

### From candidate to clinical laboratory assay.

The remaining steps in biomarker development centre on translating the candidate biomarker into a clinical-grade assay with well-defined performance characteristics. A clinical grade assay ideally uses a well-validated technology (e.g. multiplex quantitative polymerase chain reaction [qPCR] or enzyme-linked immunosorbent assay [ELISA]) that is robust to delays in sample processing (e.g. whole blood collected into RNA-stabilising tubes). By the time it reaches the clinic, the biomarker's performance using 'real-world' cohorts across multiple centres should be well known. Exactly how these later stage trials are structured will depend upon the nature of the biomarker (e.g. diagnostic (versus similar diseases or reflecting disease activity) or prognostic (of disease course untreated or of response to specific therapies)).

A subset of biomarkers with potentially great impact in rheumatology are those intended as endpoints or to identify patient subgroups in clinical trials. The increasing interest in identifying patient subgroups stems from the recognition that while many newer therapies target specific biologic pathways (e.g. infliximab and the TNF-alpha pathway in rheumatoid arthritis, sifalimumab and the type 1 interferon pathway in SLE and autoimmune myositis), the relationship between biologic pathway and disease is not always one to one. For example, a subset of patients with clinical SLE will have no identifiable dysregulation of the type 1 interferon pathway and a plausible hypothesis would be that this subset of patients will not benefit from anti-interferon therapy [24]. If this proves to be the case

then using upregulation of IFN-I genes to choose patients for anti-interferon alpha therapy should result in increased clinical response rates. This in turn should allow pre-registration clinical studies to be adequately powered with fewer patients and ultimately make costly medications more cost-effective to health-care funding bodies. It should be noted that the bar for validation of biomarkers intended for drug development is set higher, with specific regulatory processes laid down by the US Food and Drug Administration and European Medicines Agency [25,26].

Recent court cases have also raised questions about the patent status of gene-based tests. The key decision was handed down in June 2013, when the United States Supreme Court decided in *Association of Molecular Pathology v. Myriad Genetics* that genomic DNA and the information therein was patent-ineligible. The subtleties and consequences of this decision are evolving, but the additional physical processing and intellectual property inherent in a gene expression-based biomarker may leave this class of test unaffected [27,28].

## 2. Gene expression-based biomarkers in rheumatology.

### The type 1 interferon signature.

A role for type 1 interferons in SLE was first recognised in the late 1970s [29]. It has also long been known that therapy with exogenous type 1 interferons would occasionally precipitate a lupus-like syndrome. In vitro work in 2001 then provided additional evidence that interferon-alpha may contribute to the pathogenesis of SLE through its effects on the differentiation of dendritic cells [30]. So the 2003 publication of two independent microarray-based papers describing a marked upregulation of IFN-I genes ('the interferon signature') in the peripheral blood of SLE patients was not wholly unexpected. However, the striking prominence of this interferon signature, its apparent association with disease activity, and its sensitivity to high dose steroid were novel findings that raised the possibility of a much-needed objective disease activity biomarker for SLE [4,5].

Research into the interferon signature since these initial observations has proceeded along the following lines: (i) the clinical significance of the interferon signature in SLE, (ii) its translation into a clinical diagnostic test and subsequent application as a pharmacodynamic marker in trials of anti-interferon therapy, and (iii) its role in other autoimmune diseases (Table 1).

### *The clinical significance of an interferon signature in SLE.*

One of the most enticing possibilities has been that the interferon signature represents a biomarker for SLE activity. A problem for researchers has been how to test this possibility in the absence of a definitive gold standard. Generally, studies aim to show that the interferon signature correlates with one of two broadly used clinical disease activity scores: SLEDAI, or BILAG [31,32]. Both are well-validated instruments that have been shown to correlate with disease activity surrogates such as escalation of treatment and clinician's impression [33]. However, as such they incorporate clinical assumptions about the significance of inflammation in one organ system (e.g. central nervous system) versus another (e.g. skin). For this reason it may be inappropriate to expect a biologically derived marker of disease activity to follow these scores closely. Instead, it is better that a proposed activity biomarker show moderate correlation with multiple relevant parameters (e.g. number of organ systems involved, treatment escalation) than excellent correlation with one.

There are now numerous studies examining the correlation between SLE activity and the IFN-I signature, using multiple versions of the interferon signature and a number of ways of assessing disease activity (Table 1). In our experience, different versions of the interferon signature are so tightly correlated that it is reasonable to consider these studies together. Of these studies, those with a cross-sectional design show a consistent, moderate association between IFN-I gene expression and disease activity measured by SLEDAI, BILAG, number of organ systems involved, and renal involvement, among others.

Unfortunately, the two published studies that follow IFI gene expression longitudinally report that the interferon score does not track disease activity within an individual over time [34,35]. Instead, its expression in an individual appears to be largely stable when followed over a period of months. In the longer-term, the expression of the interferon signature appears to decline with age from a peak in the second and third decade [36]. Indeed, these data would suggest that the interferon signature in SLE may be better thought of a risk factor predicting more severe disease than as a biomarker of day to day activity that can be used to detect flares in the clinic.

The ability of an interferon high state to predict disease activity was also addressed by Bauer *et al* in studies of interferon inducible serum cytokines in SLE [37]. They used a protein microarray platform

to assay 160 serum cytokines in a cohort of SLE patients. In doing so, they identified a cluster of twelve cytokines that were demonstrably interferon inducible *in vitro* and correlated with the IFN-I gene expression signature. A composite chemokine score was derived from three chemokines in this cluster (CCL2, CCL19 and CXCL10) and taken forward for independent validation [38]. A high chemokine score in the validation cohort was sensitive (sensitivity 83%) but not specific (specificity 49%) for detecting active disease. However, a high chemokine score was associated with an increased hazard of 2.5 for an SLE flare in the following 12 months suggesting that this assay (and perhaps by extrapolation the IFN-I signature) may be best used to risk stratify patients in remission.

Other non-expression-based surrogate markers for the interferon signature have also been studied (i.e. Siglec-1 expression or CD64 surface expression on monocytes by flow cytometry) and in these a correlation with disease activity is also observed [39,40].

Nearly all studies of the interferon signature in SLE include a subset of patients with at least moderate upregulation of IFN-I genes despite clinically inactive disease and another of patients with clinically active disease, and no detectable interferon signature. Are there subgroups of SLE patients that preferentially exhibit an interferon signature independently of disease activity? This is a difficult question to address, because many relevant subgroups (e.g. cerebral lupus, renal involvement) also identify patients with more severe disease. However, one study found that after adjusting for SLEDAI, anti-U1RNP antibodies remained independently associated with interferon-high patients [41]. This is consistent with other univariate analyses demonstrating an association between anti-RNP, anti-Ro and anti-La autoantibodies and an interferon signature in SLE [42,43].

Evidence is accumulating that an active interferon signature may be associated with worse vascular disease in SLE, an important finding given the contribution of accelerated vascular disease to long-term mortality in this condition. Lood *et al* report that an interferon signature in platelets from SLE patients correlates strongly with a history of vascular disease [44]. Somers *et al* report an association between interferon signature and coronary calcification and carotid intima media thickness, both early subclinical markers of vascular disease [45]. Finally, single nucleotide polymorphisms in IRF8, a transcription factor of the type 1 interferon pathway, have been associated with ischaemic heart disease in SLE, albeit at sub-genomewide significance in a targeted "within cases" genetic analysis [46].

### *The interferon signature to guide anti-interferon alpha therapy.*

Recently, the focus has shifted from the interferon signature as a disease activity marker to its ability to select patients for and as a pharmacodynamic marker of anti-type 1 interferon therapy. As an adjunct to their development of sifalimumab, a human anti-interferon alpha monoclonal antibody, MedImmune have defined and validated a whole blood IFN-I signature assay. The assay output is a single continuous interferon score, based on the expression of 21-genes. The genes were chosen from those upregulated after PBMC stimulation *ex vivo* with type 1 interferon and SLE patient serum. This upregulation had to be attenuated in the presence of anti-interferon alpha and anti-interferon alpha receptor monoclonal antibodies and the genes also had to be overexpressed in the blood of SLE patients versus healthy volunteers [47].

This biomarker was successfully able to identify type 1 interferon high samples at baseline and inhibition of the signature by sifalimumab in a phase 1 randomised controlled trial [48]. As a qPCR-based assay, it exhibited high correlation with a microarray-based gold standard determined in parallel and also with interferon activity in paired skin biopsies.

The assay has subsequently been implemented in a follow-up phase 1b study of sifalimumab where it was used in block randomization of patients at baseline, and as an exploratory pharmacodynamic endpoint. This study again confirmed dose-dependent, incomplete inhibition of the IFN-I signature and found a trend to improved disease activity in the subgroup that were interferon high at baseline and treated with sifalimumab [24]. It is not surprising that inhibition was incomplete: only a subset of type 1 interferons are neutralized by the two current anti-interferon alpha antibodies, sifalimumab and rontalizumab. Indeed, the observation highlights the importance of including the interferon signature as a pharmacodynamic marker in these studies.

A 7-gene qPCR assay was developed by Genentech for their phase 1 study of rontalizumab in SLE [49]. Again they were able to identify interferon high patients (as compared with a microarray gold standard) and demonstrate dose-dependent reduction in interferon score after rontalizumab. This assay was subsequently used to classify participants into interferon high and low groups for a follow-up phase 2 study of subcutaneous or intravenous rontalizumab versus placebo in moderate to severe SLE. Although currently only published in abstract form this study concluded, paradoxically, that only patients who were interferon low at baseline had a significant response to anti-interferon therapy [50].

It is difficult to know how to interpret this in light of the sifalimumab observations above, and factors such as other clinical and laboratory differences between the interferon high and low groups, and the degree of inhibition achieved by rontalizumab in the interferon high group will have to be examined carefully when the study is ultimately published.

A separate, 13-gene, interferon assay has also been developed specifically for trials in dermatomyositis and polymyositis. Although the assay was developed from expression data generated in a longitudinal study of dermatomyositis and polymyositis patients, the 13 genes chosen are a subset of the 21-gene MedImmune assay, and there is a tight correlation between the two (r=0.82) [51]. This has since been used for a phase 1b clinical trial of sifalimumab in autoimmune myositis, with similar outcome to the SLE studies above [52].

In summary, a range of assays based on overlapping lists of IFI genes are entering use in clinical trials of anti-interferon therapy. From the initial reports, these biomarkers have highly similar properties and all have been validated to varying degrees as surrogate markers for a microarray defined interferon signature.

### *Type 1 interferon signature in diseases other than SLE.*

Since the identification of a type 1 interferon signature in SLE, similar signatures have been described for a number of other autoimmune diseases [53–55]. In a 2011 study, Higgs *et al* compared the whole blood transcriptome of patients with SLE, dermatomyositis, polymyositis, rheumatoid arthritis and systemic sclerosis [53]. In each of these diseases there was a subset of patients with a detectable interferon signature. Moreover, by comparing differentially expressed gene lists and by pathway analysis, the authors showed that the genes comprising the interferon signature in each disease were broadly similar. The proportions of polymyositis, dermatomyositis and systemic sclerosis patients with a detectable interferon signature (using a threshold determined based on healthy volunteers) were similar to that of SLE (61, 66, 68 & 73% respectively), whereas the proportion of rheumatoid arthritis patients with a detectable signature was only one third. These frequencies are similar to previous reports [5,55,56]. A substantial majority of patients with Sjogren's syndrome also demonstrate an interferon signature [57]. Other autoimmune rheumatic diseases, notably AAV, have been studied by microarray and are not associated with upregulation of type 1 interferon [58]. As with SLE, researchers

have shown an association between the interferon signature and disease activity for systemic sclerosis and polymyositis [53]. In rheumatoid arthritis, two publications confirm that an interferon signature at baseline is associated with a poor response to rituximab and in Sjogren's syndrome, the signature was shown to be associated with anti-Ro and anti-La autoantibody titres [57,59,60].

Does the presence of an interferon signature allow conclusions to be drawn regarding a common pathogenesis for these disorders? Certainly, the overlap between SLE and Sjogren's syndrome is well described and recent *in vitro* work has shown a link between anti-nucleic acid antibodies and interferon production [54,61]. Anti-nucleic antibodies (e.g. anti-Jo1, targeting a tRNA synthetase) have also been described in the inflammatory myopathies and systemic sclerosis is characterized by a range of autoantibodies against nuclear components, albeit not always nucleic acids. An interferon signature has been shown to predate the onset of clinical disease in a cohort at high risk for developing rheumatoid arthritis [62].

### *Summary – Type 1 interferon signature*

The type 1 interferon signature remains the most well-studied gene expression-based biomarker in autoimmune disease. It is associated with more severe disease in SLE, but is insensitive to short-term fluctuations in disease activity. Subgroups of patients from an expanding number of autoimmune diagnoses can be identified to have similar upregulation of interferon signaling pathways.

A number of unanswered questions remain. In particular, differences in interferon signature between disease or in individual cell types have not been studied in humans to any great extent. Given the increasing appreciation of the complexity of interferons (the type 1 interferon family comprises 16 members) and their downstream signaling pathways even small differences may be important (e.g. reflecting shifts in the balance of type 1 interferon cytokines) [63].

There are two likely applications for an interferon-based assay in clinical practice. First, given the association with disease activity in cross-sectional studies, it may prove to be useful in risk-stratifying SLE patients for planning maintenance therapy during remission. Second, if anti-interferon antibodies enter clinical use, an interferon score may be useful in identifying patients for treatment and providing pharmacodynamic correlation with treatment efficacy.

### A CD8+ T-cell prognostic signature in AAV and SLE.

In 2010 McKinney *et al* reported a CD8+ T-cell gene signature that robustly divides flaring AAV and SLE patients into good and poor prognostic groups [6]. They began with the observation that when the unfiltered transcriptome of purified peripheral blood CD8+ T-cells was subjected to unsupervised clustering, samples separated neatly into two statistically distinct subgroups. Subgroup membership was most highly correlated with number of flares and time to flare over a median follow-up of three years. A shortlist of the most influential genes defining these subgroups was able to recapitulate this prognostic classification in a validation cohort of 27 flaring AAV patients and in a third cohort of 26 flaring SLE patients (Figure 2).

A follow-up study from the same laboratory reports that a highly similar CD8+ T-cell signature classifies patients with active inflammatory bowel disease into good and poor prognostic subgroups [64]. Such a consistent observation across four distinct autoimmune and autoinflammatory conditions suggests the signature reflects fundamental differences in T-cell biology. Genes encoding T-cell receptor and IL-7 receptor signaling pathway components were enriched in the poor prognosis subgroup, hinting at a role for T-cell survival and the development of T-cell memory.

The ability to classify patients at presentation by prognosis using a small number of CD8+ T-cell expressed genes raises the real possibility of personalizing treatment. Patients in the poor prognosis group may benefit from an intensified maintenance immunosuppression regimen, while those with a good prognosis could potentially be weaned off maintenance therapy more rapidly. However, a number of hurdles stand between this biomarker and its clinical application. It must be transformed from a technically complex marker based on fresh, separated leukocytes into a simplified, robust clinical assay. Alongside this, a strategy for classifying new samples as they arise needs to be developed. The new assay will then require validation both against the original CD8+ T-cell based classification and against clinical prognosis. Other open questions, such as the effect of treatment on classification, and the longitudinal stability of prognostic assignments are also important given the broad range of clinical scenarios in which this biomarker would be of potential use.

### Peripheral blood transcriptional modules and SLE.

Chaussabel and colleagues have used a module-based approach to build a potential disease activity based biomarker for paediatric SLE [7]. They subjected PBMC-derived gene expression data from 239 patients with a broad range of immune-related diagnoses including sepsis, organ transplantation and autoimmune disease to repeated rounds of gene clustering to yield 28 modules. Each module represents a cluster of genes that are tightly correlated across at least six out of the eight diagnoses in the dataset. Literature mining allowed functional annotations to be attached to most, but not all, of these modules. The output from an individual module - a single value reflecting whether genes in that module are highly or lowly-expressed in a given sample – is combined with that of the other modules to generate an 'at a glance' summary of the PBMC transcriptome for a given patient at a given point in time.

The validity of this approach was established by demonstrating that type 1 interferon, 'granulopoiesis' and plasma cell modules were upregulated in SLE in a novel SLE cohort. Summary measures of these and 8 additional modules that were also correlated with SLE were then examined for their association with disease activity and and for confounding by treatment. Concomitant steroid or cytotoxic therapy appeared not to affect this modular signature of SLE, but certain modules ('interferon', 'immature granulocytes', 'erythrocytes', 'T-cell lymphopaenia' and 'ribosomal down-regulation') did correlate with disease activity. A quantitative PCR assay using two genes from each of these five modules was also correlated with activity (spearman rho=0.63 for an independent SLE cohort). A composite 'u-score' was then used to summarize gene expression in these modules. This score has the advantage of being non-parametric and scale independent, but as at it relies on a given sample's gene expression in relation to others, it may not translate seamlessly to a clinical setting. Nonetheless, the u-score mirrored changes in SLEDAI over time (2-4 time points) for most patients assessed and whilst individual gene expression correlated poorly across platforms, module level expression was at least moderately correlated when two markedly different array technologies were applied to the same samples.

This work is ongoing: a re-analysis has replaced the initial 28 transcriptional modules with a set of 260, derived from an expanded set of now whole blood gene expression data obtained from healthy volunteers and patients with infection and immune-related conditions. Of relevance to SLE, the interferon transcriptional module was split into three, each with distinct properties. In a cohort of SLE patients, the expression of two of these modules varied over time, and one showed a modest correlation with a number of indices of disease activity. The authors show differences in the expression of these

modules in patients receiving exogenous interferon alpha and beta and hypothesise that these differences are due to differing sensitivities of the three modules to interferon-alpha, beta and gamma [65].

Overall, this implementation of disease-independent modules of correlated gene expression represents an interesting and innovative approach to biomarker development. Key advantages include the ability to combine multiple transcriptionally independent modules to improve the correlation of individual modules with relevant clinical parameters, as well as its comparatively straightforward translation to the clinic. However, even in the updated analysis many of these modules seem to detect changes in the relative proportions of leukocyte subsets within a sample, rather than cell-level alterations in gene expression. While this doesn't affect the validity of the approach, it complicates biologic interpretation and raises the question as to whether detailed immunophenotyping may be an alternative approach that is more amenable to widespread implementation.

## Other notable gene expression studies and signatures.

### *The granulopoiesis signature of SLE*

The first microarray studies of SLE patients found an overrepresentation of granulocyte-specific genes in peripheral blood mononuclear cell samples. Flow cytometry subsequently confirmed that this was due to the presence of immature granulocytes retained within the PBMC layer, as had been reported previously [5,66]. This granulopoiesis signature correlated with disease activity at least as well as the interferon signature, but unlike the interferon signature has not been developed further as a potential biomarker. Instead, the observation has stimulated a substantial body of research into the role of neutrophils in SLE, uncovering substantial interactions between neutrophils and interferon-producing plasmacytoid dendritic cells [67].

### *Gene expression profiling in ankylosing spondylitis*

The two major studies of peripheral blood gene expression in ankylosing spondylitis (AS) have conflicting findings. The first used PBMC samples from a cohort of 18 AS patients, ten of whom were receiving immunosuppression (anti-TNF therapy and/or corticosteroid), and age- and gender matched healthy volunteers [68]. Differential expression analysis identified predominantly downregulated immune-associated genes, and four of these (NR4A2, TNFAIP3, CD69 and RORA) were chosen for qPCR validation in a second cohort of 35 AS patients. NR4A2, TNFAIP3 and CD39 were also significantly underexpressed in the validation cohort, with an area under receiver operator curve of 77% for discriminating AS from healthy volunteers. However, the second study examined the whole blood gene expression of 16 patients with untreated, active AS, reporting significant upregulation of TLR4 and TLR5 as well as other genes in related pro-inflammatory pathways, confirmed in a second cohort [69]. Although identified by the first study, NR4A2, TNFAIP3 and CD69 were not differentially expressed in this analysis. Differences in sample type (whole blood as compared with PBMC) may explain this, but the results of the first study may also be confounded by immunosuppression. Downregulation of TNFAIP3 is reported as predictive of response to anti-TNF therapy, and reduced CD69 expression in lymphocyte subsets has been reported in cells treated *in vitro* with etanercept [70,71].

## Conclusion

The process of discovering gene expression-based biomarkers is painstaking, and the process of turning these candidate biomarkers into a practical and useful clinical test even more so. Biology does not always line up neatly with clinical parameters. Peripheral blood is a useful but imperfect mirror for disease processes occurring in tissues. Processing and technical artefacts are difficult to avoid.

Despite this, after a decade of gene expression-based biomarker research in rheumatology a number of promising candidates are moving forward. The need for a pharmacodynamic marker for anti-interferon therapy has driven forward development of the interferon signature into well-validated qPCR-based assays that are now integral to a number of clinical trials. Our current understanding of the interferon signature in autoimmune disease suggests that although it is a risk factor for more active disease, it

does not reflect short term changes in disease activity and perhaps will be most clinically relevant in stratifying patients in remission for maintenance therapy.

Another candidate biomarker of prognosis is the CD8+ T-cell marker described by McKinney *et al* [6,64]. Although in its current form the marker is technically complex and more suited for a research setting, it offers the prospect of tailoring a patient's treatment to their prognosis on presentation.

Chaussabel *et al* have developed a flexible framework for biomarker development in autoimmune disease using whole blood samples [7]. By combining multiple independent modules of expression they have developed a biomarker of SLE disease activity with impressive performance in a small validation cohort. A larger validation study will be important to determining its clinical role.

Looking back over the last decade of gene expression-based biomarker research, an approach characterized by individual centres using small cohorts and unique protocols has been a significant brake on the pace and scope of biomarker discovery. Multicentre collaborations are already starting to emerge (e.g. the FOCIS Human Immunophenotyping Consortium, working to standardize immunophenotyping panels and procedures between centres) and developing these collaborations further using common methodology optimized for emerging high throughput technology is likely to be the key to making the next decade of research even more productive than the last.

## References


1. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. J. Clin. Oncol. 2006;24:3726–34.

2. Pham MX, Teuteberg JJ, Kfoury AG, Starling RC, Deng MC, Cappola TP, et al. Gene-expression profiling for rejection surveillance after cardiac transplantation. N. Engl. J. Med. 2010;362:1890–900.

3. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J. Natl. Cancer Inst. 2007;99:147–57.

4. Baechler EC, Batliwalla FM, Karypis G, Gaffney PM, Ortmann WA, Espe KJ, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. Proc. Natl. Acad. Sci. U. S. A. 2003;100:2610–5.

5. Bennett L, Palucka a K, Arce E, Cantrell V, Borvak J, Banchereau J, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J. Exp. Med. 2003;197:711–23.

6. McKinney EF, Lyons PA, Carr EJ, Hollis JL, Jayne DRW, Willcocks LC, et al. A CD8+ T cell transcription signature predicts prognosis in autoimmune disease. Nat. Med. 2010;16:586–91.

7. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity 2008;29:150–64.

8. Atkinson J, Colburn W a., DeGruttola VG, DeMets DL, Downing GJ, Hoth DF, et al. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin. Pharmacol. Ther. 2001;69:89–95.

9. Prensner JR, Chinnaiyan AM, Srivastava S. Systematic, evidence-based discovery of biomarkers at the NCI. Clin. Exp. Metastasis 2012;29:645–52.

10. Newell KA, Asare A, Kirk AD, Gisler TD, Bourcier K, Suthanthiran M, et al. Identification of a B cell signature associated with renal transplant tolerance in humans. J. Clin. Invest. 2010;120:1836–1847.

11. Chai V, Vassilakos A, Lee Y, Wright J a, Young AH. Optimization of the PAXgene blood RNA extraction system for gene expression analysis of clinical samples. J. Clin. Lab. Anal. 2005;19:182–8.

12. Lyons PA, McKinney EF, Rayner TF, Hatton A, Woffendin HB, Koukoulaki M, et al. Novel expression signatures identified by transcriptional analysis of separated leucocyte subsets in systemic lupus erythematosus and vasculitis. Ann. Rheum. Dis. 2010;69:1208–13.

13. Hanley MB, Lomas W, Mittar D, Maino V, Park E. Detection of low abundance RNA molecules in individual cells by flow cytometry. PLoS One 2013;8:e57002.

14. Robinson MD, Speed TP. A comparison of Affymetrix gene expression arrays. BMC Bioinformatics 2007;8:449.

15. Zakharkin SO, Kim K, Mehta T, Chen L, Barnes S, Scheirer KE, et al. Sources of variation in Affymetrix microarray experiments. BMC Bioinformatics 2005;6:214.

16. Akey JM, Biswas S, Leek JT, Storey JD. On the design and analysis of gene expression studies in human populations. Nat. Genet. 2007;39:807–8

17. Ransohoff DF, Gourlay ML. Sources of bias in specimens for research about molecular markers for cancer. J. Clin. Oncol. 2010;28:698–704.

18. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531–7.

19. Ressom HW, Varghese RS, Zhang Z, Xuan J, Clarke R. Classification algorithms for phenotype prediction in genomics and proteomics. Front. Biosci. 2008;13:691–708.

20. Liu Y, Hayes DN, Nobel A, Marron JS. Statistical Significance of Clustering for High-Dimension, Low–Sample Size Data. J. Am. Stat. Assoc. 2008;103:1281–93.

21. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. Nat. Rev. Cancer 2005;5:142–9.

22. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J. Natl. Cancer Inst. 2003;95:14–8.

23. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nat. Rev. Cancer 2004;4:309–14.

24. Petri M, Wallace DJ, Spindler A, Chindalore V, Kalunian K, Mysler E, et al. Sifalimumab, a human anti-interferon-α monoclonal antibody, in systemic lupus erythematosus: a phase I randomized, controlled, dose-escalation study. Arthritis Rheum. 2013;65:1011–21.

25. Scherf U, Becker R, Chan M, Hojvat S. Approval of novel biomarkers: FDA's perspective and major requests. Scand. J. Clin. Lab. Invest. Suppl. 2010;242:96–102.

26. Goodsaid F, Papaluca M. Evolution of biomarker qualification at the health authorities. Nat. Biotechnol. 2010;28:441–3.

27. Rai AK, Cook-Deegan R. Genetics. Moving beyond "isolated" gene patents. Science 2013;341:137–8.

28. Kesselheim AS, Cook-Deegan RM, Winickoff DE, Mello MM. Gene patenting--the Supreme Court finally speaks. N. Engl. J. Med. 2013;369:869–75.

29. Hooks JJ, Moutsopoulos HM, Geis SA, Stahl NI, Decker JL, Notkins AL. Immune interferon in the circulation of patients with autoimmune disease. N. Engl. J. Med. 1979;301:5–8.

30. Blanco P, Palucka a K, Gill M, Pascual V, Banchereau J. Induction of dendritic cell differentiation by IFN-alpha in systemic lupus erythematosus. Science 2001;294:1540–3.

31. Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. J. Rheumatol. 2002;29:288–91.

32. Isenberg D a, Rahman A, Allen E, Farewell V, Akil M, Bruce IN, et al. BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. Rheumatology (Oxford). 2005;44:902–6.

33. Romero-Diaz J, Isenberg D, Ramsey-Goldman R. Measures of adult systemic lupus erythematosus. Arthritis Care Res. (Hoboken). 2011;63 Suppl 1:S37–46.

34. Petri M, Singh S, Tesfasyone H, Dedrick R, Fry K, Lal P, et al. Longitudinal expression of type I interferon responsive genes in systemic lupus erythematosus. Lupus 2009;18:980–9.

35. Landolt-Marticorena C, Bonventi G, Lubovich A, Ferguson C, Unnithan T, Su J, et al. Lack of association between the interferon-alpha signature and longitudinal changes in disease activity in systemic lupus erythematosus. Ann. Rheum. Dis. 2009;68:1440–6.

36. Niewold TB, Adler JE, Glenn SB, Lehman TJ a, Harley JB, Crow MK. Age- and sex-related patterns of serum interferon-alpha activity in lupus families. Arthritis Rheum. 2008;58:2113–9.

37. Bauer JW, Baechler EC, Petri M, Batliwalla FM, Crawford D, Ortmann W a., et al. Elevated serum levels of interferon-regulated chemokines are biomarkers for active human systemic lupus erythematosus. PLoS Med. 2006;3:e491.

38. Bauer JW, Petri M, Batliwalla FM, Koeuth T, Wilson J, Slattery C, et al. Interferon-regulated chemokines as biomarkers of systemic lupus erythematosus disease activity: a validation study. Arthritis Rheum. 2009;60:3098–107.

39. Biesen R, Demir C, Barkhudarova F, Grün JR, Steinbrich-Zöllner M, Backhaus M, et al. Sialic acid-binding Ig-like lectin 1 expression in inflammatory and resident monocytes is a potential biomarker for monitoring disease activity and success of therapy in systemic lupus erythematosus. Arthritis Rheum. 2008;58:1136–45.

40. Li Y, Lee PY, Kellner ES, Paulus M, Switanek J, Xu Y, et al. Monocyte surface expression of Fcγ receptor RI (CD64), a biomarker reflecting type-I interferon levels in systemic lupus erythematosus. Arth. Res Ther. 2010;12:R90.

41. Nikpour M, Dempsey a a, Urowitz MB, Gladman DD, Barnes D a. Association of a gene expression profile from whole blood with disease activity in systemic lupus erythaematosus. Ann. Rheum. Dis. 2008;67:1069–75.

42. Kirou K a, Lee C, George S, Louca K, Peterson MGE, Crow MK. Activation of the interferon-alpha pathway identifies a subgroup of systemic lupus erythematosus patients with distinct serologic features and active disease. Arthritis Rheum. 2005;52:1491–503.

43. Zhuang H, Narain S, Sobel E, Lee PY, Nacionales DC, Kelly KM, et al. Association of anti-nucleoprotein autoantibodies with upregulation of Type I interferon-inducible gene transcripts and dendritic cell maturation in systemic lupus erythematosus. Clin. Immunol. 2005;117:238–50.

44. Lood C, Amisten S, Gullstrand B, Jönsen A, Allhorn M, Truedsson L, et al. Platelet transcriptional profile and protein expression in patients with systemic lupus erythematosus: up-regulation of the type I interferon system is strongly associated with vascular disease. Blood 2010;116:1951–7.

45. Somers EC, Zhao W, Lewis EE, Wang L, Wing JJ, Sundaram B, et al. Type I interferons are associated with subclinical markers of cardiovascular disease in a cohort of systemic lupus erythematosus patients. PLoS One 2012;7:e37000.

46. Leonard D, Svenungsson E, Sandling JK, Berggren O, Jönsen A, Bengtsson C, et al. Coronary heart disease in systemic lupus erythematosus is associated with interferon regulatory factor-8 gene variants. Circ. Cardiovasc. Genet. 2013;6:255–63.

47. Yao Y, Higgs BW, Morehouse C, de Los Reyes M, Trigona W, Brohawn P, et al. Development of Potential Pharmacodynamic and Diagnostic Markers for Anti-IFN-α Monoclonal Antibody Trials in Systemic Lupus Erythematosus. Hum. Genomics Proteomics 2009. doi:10.4061/2009/374312.

48. Yao Y, Richman L, Higgs BW, Morehouse CA, de los Reyes M, Brohawn P, et al. Neutralization of interferon-alpha/beta-inducible genes and downstream effect in a phase I trial of an anti-interferon-alpha monoclonal antibody in systemic lupus erythematosus. Arthritis Rheum. 2009;60:1785–96.

49. McBride JM, Jiang J, Abbas AR, Morimoto A, Li J, Maciuca R, et al. Safety and pharmacodynamic results of rontalizumab in a phase i, placebo controlled, double blind, dose escalation study in systemic lupus erythematosus. Arthritis Rheum. 2012;64:3666–76.

50. Kalunian K, Merril JT, Maciuca R, Ouyang W, McBride JM, Townsend M, et al. Efficacy And Safety Of Rontalizumab (Anti-Interferon Alpha) In SLE Subjects With Restricted Immunosuppressant Use: Results Of A Randomized, Double-Blind, Placebo-Controlled Phase 2 Study.[abstract]. Arthritis Rheum. 2012;64 Suppl 1:2622.

51. Greenberg SA, Higgs BW, Morehouse C, Walsh RJ, Kong SW, Brohawn P, et al. Relationship between disease activity and type 1 interferon- and other cytokine-inducible gene expression in blood in dermatomyositis and polymyositis. Genes Immun. 2012;13:207–13.

52. Higgs BW, Zhu W, Morehouse C, White WI, Brohawn P, Guo X, et al. A phase 1b clinical trial evaluating sifalimumab, an anti-IFN-α monoclonal antibody, shows target neutralisation of a type I IFN signature in blood of dermatomyositis and polymyositis patients. Ann. Rheum. Dis. 2014;73:256–62.

53. Higgs BW, Liu Z, White B, Zhu W, White WI, Morehouse C, et al. Patients with systemic lupus erythematosus , myositis , rheumatoid arthritis and scleroderma share activation of a common type I interferon pathway. Ann. Rheum. Dis. 2011;70:2029–36.

54. Manoussakis MN, Georgopoulou C, Zintzaras E, Spyropoulou M, Stavropoulou A, Skopouli FN, et al. Sjögren's syndrome associated with systemic lupus erythematosus: clinical and laboratory profiles and comparison with primary Sjögren's syndrome. Arthritis Rheum. 2004;50:882–91.

55. Van der Pouw Kraan TCTM, Wijbrandts C a, van Baarsen LGM, Voskuyl a E, Rustenburg F, Baggen JM, et al. Rheumatoid arthritis subtypes identified by genomic profiling of peripheral blood cells: assignment of a type I interferon signature in a subpopulation of patients. Ann. Rheum. Dis. 2007;66:1008–14.

56. Walsh RJ, Kong SW, Yao Y, Jallal B, Kiener PA, Pinkus JL, et al. Type I interferon-inducible gene expression in blood is present and reflects disease activity in dermatomyositis and polymyositis. Arthritis Rheum. 2007;56:3784–92.

57. Emamian ES, Leon JM, Lessard CJ, Grandits M, Baechler EC, Gaffney PM, et al. Peripheral blood gene expression profiling in Sjögren's syndrome. Genes Immun. 2009;10:285–96.

58. Alcorta DA, Barnes DA, Dooley MA, Sullivan P, Jonas B, Liu Y, et al. Leukocyte gene expression signatures in antineutrophil cytoplasmic autoantibody and lupus glomerulonephritis. Kidney Int. 2007;72:853–64.

59. Vosslamber S, Raterman HG, van der Pouw Kraan TCTM, Schreurs MWJ, von Blomberg BME, Nurmohamed MT, et al. Pharmacological induction of interferon type I activity following treatment with rituximab determines clinical response in rheumatoid arthritis. Ann. Rheum. Dis. 2011;70:1153–9.

60. Raterman HG, Vosslamber S, de Ridder S, Nurmohamed MT, Lems WF, Boers M, et al. The interferon type I signature towards prediction of non-response to rituximab in rheumatoid arthritis patients. Arthritis Res. Ther. 2012;14:R95.

61. Garcia-Romo GS, Caielli S, Vega B, Connolly J, Allantaz F, Xu Z, et al. Netting neutrophils are major inducers of type I IFN production in pediatric systemic lupus erythematosus. Sci. Transl. Med. 2011;3:73ra20.

62. Lübbers J, Brink M, van de Stadt LA, Vosslamber S, Wesseling JG, van Schaardenburg D, et al. The type I IFN signature as a biomarker of preclinical rheumatoid arthritis. Ann. Rheum. Dis. 2013;72:776–80.

63. González-Navajas JM, Lee J, David M, Raz E. Immunomodulatory functions of type I interferons. Nat. Rev. Immunol. 2012;12:125–35.

64. Lee JC, Lyons PA, McKinney EF, Sowerby JM, Carr EJ, Bredin F, et al. Gene expression profiling of CD8+ T cells predicts prognosis in patients with Crohn disease and ulcerative colitis. J. Clin. Invest. 2011;121:4170–9.

65. Chiche L, Jourde-Chiche N, Whalen E, Presnell S, Gersuk V, Dang K, et al. Modular transcriptional repertoire analyses of adults with systemic lupus erythematosus reveal distinct type I and type II interferon signatures. Arthritis Rheumatol. (Hoboken, N.J.) 2014;66:1583–95.

66. Hacbarth E, Kajdacsy-Balla A. Low density neutrophils in patients with systemic lupus erythematosus, rheumatoid arthritis, and acute rheumatic fever. Arthritis Rheum. 1986;29:1334–42.

67. Lande R, Ganguly D, Facchinetti V, Frasca L, Conrad C, Gregorio J, et al. Neutrophils activate plasmacytoid dendritic cells by releasing self-DNA-peptide complexes in systemic lupus erythematosus. Sci. Transl. Med. 2011;3:73ra19.

68. Duan R, Leo P, Bradbury L, Brown M a, Thomas G. Gene expression profiling reveals a downregulation in immune-associated genes in patients with AS. Ann. Rheum. Dis. 2010;69:1724–9.

69. Assassi S, Reveille JD, Arnett FC, Weisman MH, Ward MM, Agarwal SK, et al. Whole-blood gene expression profiling in ankylosing spondylitis shows upregulation of toll-like receptor 4 and 5. J. Rheumatol. 2011;38:87–98.

70. Koczan D, Drynda S, Hecker M, Drynda A, Guthke R, Kekow J, et al. Molecular discrimination of responders and nonresponders to anti-TNF alpha therapy in rheumatoid arthritis by etanercept. Arthritis Res. Ther. 2008;10:R50.

71. Thaher F, Plankenhorn S, Klein R. Differential effects of the tumor necrosis factor alpha-blocker infliximab and etanercept on immunocompetent cells in vitro. Int. Immunopharmacol. 2011;11:1724–31.

## Conflict of interest statement:

S.M.F.'s research is partly funded by GSK. K.G.C.S. receives consulting fees from MedImmune and UCB. P.A.L, E.F.M and K.G.C.S. are named on a patent application covering the 8.1/8.2 biomarker.

## Acknowledgements:

The authors acknowledge support from the Cambridge Biomedical Research Centre. This work was supported by the Wellcome Trust [083650/Z/07/Z, 094227/Z/10/Z] and MRC Programme grant [MR/L019027/1]. S.M.F holds a Translational Medicine and Therapeutics PhD Studentship part-funded by the Wellcome Trust and GlaxoSmithKline. The Cambridge Institute for Medical Research is in receipt of Wellcome Trust Strategic Award 079895.

## Figures

**Figure 1.** An overview of the biomarker development process, highlighting important design considerations at each step.

**Figure 2.** Kaplan-Meier plots showing shorter time to first flare in poor prognosis '8.1' SLE group (left) and poor prognosis '8.1' ANCA-associated vasculitis group (right). Significance was measured with the log-rank test. Modified from McKinney *et al,* Nat. Med. 2010; 16(5):586–91. [6]

## Tables

**Table 1.** Summary information for the primary research articles describing interferon-based biomarkers referenced in this review.

**Table 1.** Summary information for the primary research articles describing interferon-based biomarkers referenced in this review.

| PMID | Authors | Year | Disease | Sample | Technology | Disease population | Relevance |
|---|---|---|---|---|---|---|---|
| **Gene expression-based interferon biomarkers in SLE** | | | | | | | |
| PMC151388 | Baechler E *et al* | 2003 | SLE | PBMC | Microarray (Affy U95A) | 48 cross-sect SLE | Demonstrated IFN-I signature and that it correlates with clinical activity. |
| PMC2193846 | Bennett L *et al* | 2003 | SLE | PBMC | Microarray (Affy U95A) | 30 paed. cross-sect SLE | Derived 10 gene IFN-I signature correlates with clinical activity; extinguished by MPL |
| 18063674 | Nikpour M *et al* | 2008 | SLE | WB | Microarray (Custom) | 269 cross-sect SLE | In a multivariate model, show that IFN-I score associated with SLEDAI, age and U1RNP Ab. |
| 18772188 | Landolt-Marticorena C *et al* | 2009 | SLE | WB | 5-gene qPCR | 94 SLE, longitudinal | Negative study: Change in interferon score did not mirror change in activity between two time-points 3-12 months apart. |
| PMC2950308 | Yao Y *et al* | 2009 | SLE | WB | Microarray (Affy U133A) | 41 SLE; 54 SLE for valid. | Derived a 21-gene interferon assay for subsequent use in clinical trials. |
| 19762399 | Petri M *et al* | 2009 | SLE | WB | Microarray (Agilent) | 66 cross-sect SLE | Despite showing assoc. with disease activity, a 9 patient longitudinal component showed no association with flare. |
| 20538795 | Lood C *et al* | 2010 | SLE | Platelets | Microarray (Affy U133A-2) | 10 SLE; 69 SLE valid. | Demonstrated an IFN-I signature in platelets, which was associated with vascular disease. |
| **Non-gene expression based biomarkers for the interferon signature in SLE** | | | | | | | |
| PMC1702557 | Bauer J *et al* | 2006 | SLE | Serum | Custom protein microarray | 15 IFN-I high SLE; 15 IFN-I low SLE. | Identified a subset of serum proteins that are significantly correlated with IFN-I genes and disease activity. |
| 18383365 | Biesen *et al* | 2008 | SLE | Monocytes | Siglec-1 on monocytes (FACS) | 9 active SLE; 52 SLE for validation | Demonstrated Siglec-1 on monocytes (FACS) correlates with microarray IFN-I sig. and disease activity. |
| PMC2842939 | Bauer J *et al* | 2009 | SLE | Serum | CCL2, CCL19 & CXCL10 | 267 SLE patients, longitudinal subset. | Demonstrated a composite chemokine score correlates with current SLE activity and can predicts flare over time. |
| PMC2911874 | Li Y *et al* | 2010 | SLE | WB | FcGRI (CD64) by qPCR/FACS | 108 cross-sect SLE | Demonstrated CD64 upregulated in SLE, and a modest correlation with IFN-I score. |
| **Interferon-based biomarkers in clinical studies of anti-interferon therapy** | | | | | | | |

| PMID | Authors | Year | Disease | Sample | Technology | Disease population | Relevance |
|---|---|---|---|---|---|---|---|
| 19479852 | Yao Y *et al* | 2009 | SLE | WB | Microarray (Affy U133+ 2) | 62 SLE randomized 2:1 to anti-IFNa. | 21-gene IFN-I signature in WB was partially inhibited by single dose sifalimumab |
| 22833362 | McBride *et al* | 2012 | SLE | WB | 7-gene qPCR | 60 SLE enrolled in multiple groups. | Partial dose-dependent inhibition of 7-gene IFN-I signature after rontalizumab. |
| 23434567 | Higgs B *et al* | 2013 | DM/PM | WB | Microarray (Affy) | 26 DM; 25 PM, allocated to | Partial suppression of IFN-I genes by a single dose of sifalimumab. |
| PMC3654174 | Petri M *et al* | 2013 | SLE | WB | 21-gene qPCR | 161 SLE randomized to 4 doses anti-IFNa | Partial suppression of IFN-I genes by a single dose of sifalimumab. |
| **The interferon signature in diseases other than SLE** | | | | | | | |
| PMC2443782 | Walsh R *et al* | 2007 | DM/PM | PBMC | Microarray (Affy U133A 2) | 12 DM; 11 PM. | Demonstrates an IFN-I gene signature in a cohort with inflammatory myopathy. |
| PMC3273959 | Emamian *et al* | 2009 | Sjog | PBMC | Microarray (Affy U95A2) | 21 cross-sectional SS; 17 in valid. | Demonstrates an IFN-I signature correlates with anti-Ro/anti-La titres in SS cohort. |
| 22043277 | Reynier *et al* | 2011 | RA | WB | Microarray (Affy U133+ 2) | 102 cross-sectional RA patients | Demonstrates an IFN-I signature in approximately 1/3 of a RA cohort. |
| 21444302 | Vosslamber S *et al* | 2011 | RA | WB | Microarray (Illumina) | 13 RA patients; 9 patient valid. | In comparison with pre-dose levels, induction of IFN-I genes at 3 months post-rituximab was associated with response to therapy in RA. |
| 21803750 | Higgs B *et al* | 2011 | SLE/DM/PM/ RA/SSc | WB | Microarray (Affy U133A) | 262 SLE, 44 DM, 33 PM, 28 SSc 89 RA | Large study enabling direct comparison of the interferon signature in different diseases. |
| PMC3446469 | Raterman H *et al* | 2012 | RA | WB | Microarray (Illumina) | 14 RA patients pre-rituximab. | IFN-I gene expression pre-dose correlated with response to rituximab in RA. |
| 21881594 | Greenberg S *et al* | 2012 | DM/PM | WB | Microarray (Affy U133A) | 15 DM, 9 PM, longitudinal. | Demonstrates IFN-I score in DM correlates disease activity at baseline and over time. |
| 23434571 | Lubbers J *et al* | 2013 | RA | WB | 7-gene qPCR | 115 seropositive at risk patients & 25 presymptomatic RA | Demonstrates in 2 cohorts that IFN-I high patients were more likely to progress to develop RA. |

PMID Pubmed Central or Pubmed identifier. RA Rheumatoid Arthritis. Sjog Sjogren's Syndrome. SSc Systemic Sclerosis. DM Dermatomyositis. PM. Polymyositis. SLE Systemic Lupus Erythematosus; IFN-I Type 1 interferon inducible. WB whole blood. PBMC peripheral blood mononuclear cells

**Figure 1.** An overview of the biomarker development process, highlighting important design considerations at each step.

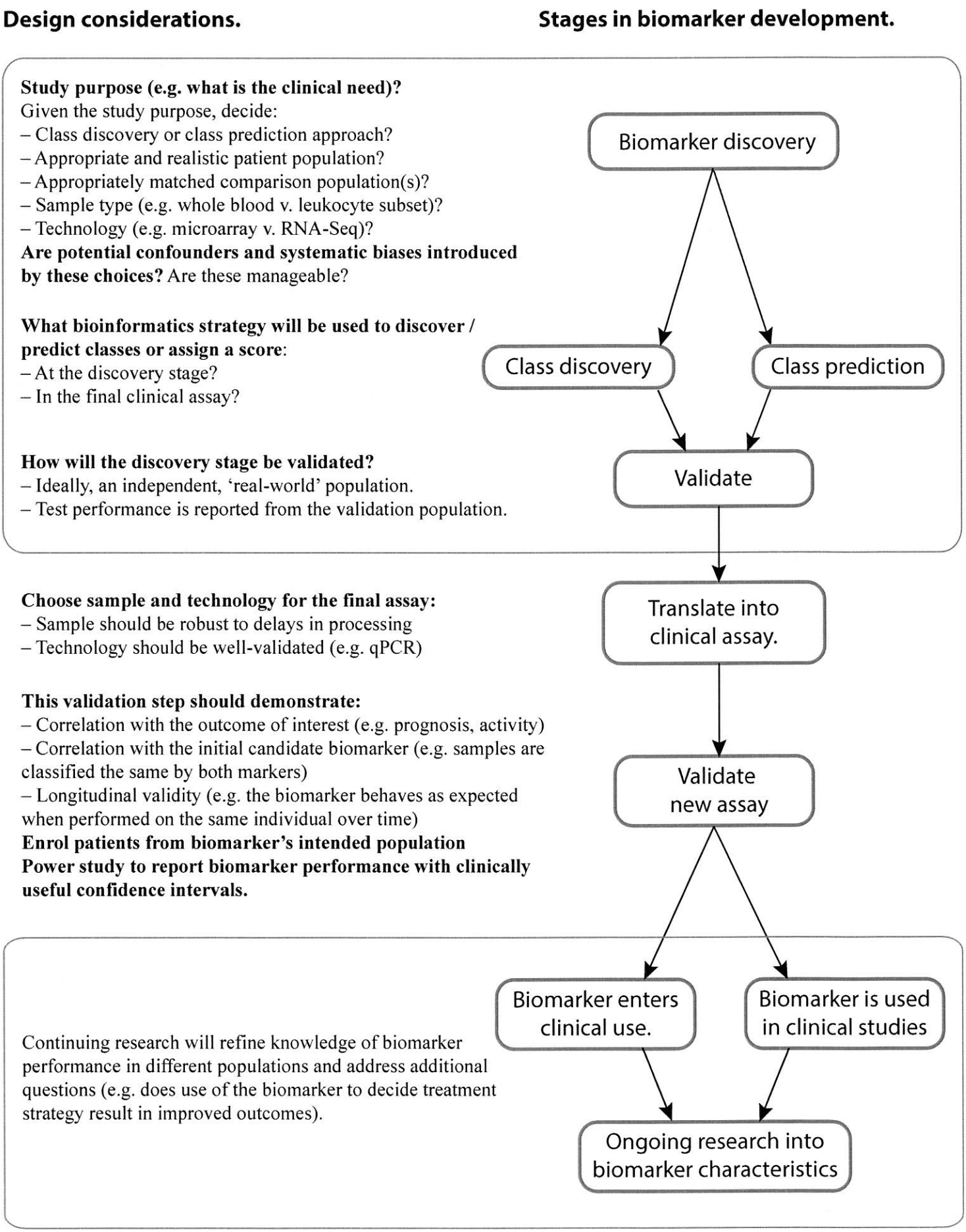

**Figure 2.** Kaplan-Meier plots showing shorter time to first flare in poor prognosis '8.1' SLE group (left) and poor prognosis '8.1' ANCA-associated vasculitis group (right). Significance was measured with the log-rank test. Modified from McKinney et al, Nat. Med. 2010; 16(5):586–91. [6]