Genome **Medicine**

RESEARCH HIGHLIGHT

Combining DNA-microarray data in systemic lupus erythematosus

Cornelis L Verweij* and Saskia Vosslamber

See related research by Arasappan *et al.*, http://www.biomedcentral.com/1741-7015/9/65

Abstract

Systemic lupus erythematosus is a systemic, heterogeneous autoimmune disease. Understanding of its molecular complexity is incomplete and there is a need to identify new therapeutic targets and to optimize criteria for its diagnosis, assessment and prognosis. Recently, Arasappan and colleagues have described a new meta-analysis method that enables data analysis across different DNA-microarray datasets to identify genes and processes relevant to systemic lupus erythematosus. Their study provides a simple and valuable meta-analysis method for the selection of biomarkers and pathways in disease.

Systemic lupus erythematosus (SLE) is a heterogeneous, multi-system autoimmune disease that presents with a wide range of clinical and laboratory abnormalities. Challenges in the clinical management of SLE include the identification of new and relevant therapeutic targets and of specific biomarkers that can be used to optimize diagnosis, assessment of disease activity (severity of disease) and prediction of flares (periodic worsening of symptoms). DNA microarray technology allows a hypothesis-free method to comprehensively identify the genes and biological pathways that are associated with clinically defined conditions. For SLE, the blood is an easily accessible compartment for monitoring immune pathophysiology by microarray analysis. Accordingly, several publications have reported microarray studies of peripheral blood cells to identify gene expression signatures in SLE. These studies mainly confirm and extend the central role of the type I interferons in SLE [1].

*Correspondence: c.verweij@vumc.nl
Department of Pathology, Section of Inflammatory Disease Profiling, CCA260,
VU University Medical Center, 1007 MB Amsterdam, the Netherlands

BioMed Central

In the May issue of *BMC Medicine*, Arasappan and colleagues [1] describe a new meta-analysis method that allows analysis across different DNA-microarray datasets to identify genes and processes relevant to SLE.

Meta-analysis across DNA-microarray datasets

Because of the complexity of microarray technology and frequently sub-optimally powered studies, verification of results is an essential step in microarray analysis. Therefore, combining analyses from different studies is important for increasing power, reliability and validation. However, several important challenges need to be considered when integrating microarray datasets for meta-analysis. First, sample collection, annotation, processing and preparation need to be performed according to quality controlled and compatible, preferably standardized, procedures. There is considerable inter-individual variability in the transcriptomes of SLE patients, which is inherent to the heterogeneity of the disease and affects analyses. Second, good laboratory techniques for data acquisition need to be used. Third, appropriate and properly used data analysis practices are required. In order to establish quality criteria and allow comparisons across independent datasets, standards for microarray experiments and data analysis were created [2]. Recent reports from the Microarray Quality Control (MAQC) consortium confirm that microarray technology is robust and should be able to reliably reveal differentially expressed genes across samples using different datasets [3].

Several approaches have been used for meta-analysis of microarray data to enable comparative analyses across multiple datasets, to minimize noise and to generate multivariate metrics for clinical use. Initial studies compared statistical measures of differentially expressed genes for each dataset to classify samples. Others revealed that the concordance between datasets improved markedly when the quantity of differential expression was used for gene selection, rather than the statistical significance [4]. Alternatively, the inter-gene correlations between datasets and a 'meta-review' method for ranking genes using the genuine published

evidence for each study have been applied. All these approaches are gene-based, that is, they aimed to identify commonly expressed genes between studies.

Application of DNA-microarray meta-analysis methods to SLE

Arasappan and colleagues [1] present a new strategy for microarray meta-analysis that is based on the identification of pathways that are coordinately expressed in multiple disease datasets. They used as input microarray datasets of peripheral blood mononuclear cell samples obtained from SLE patients, and healthy controls derived from four different studies, of which two involved only children and the other two only adults. Transcriptional profiles were generated and low stringency and fold change cut-offs were applied to select differentially expressed genes. For each dataset, Ingenuity Pathway Analysis was used to identify biological pathways that were differentially expressed between SLE patients and controls. Validation with the leave-one-out permutations method revealed three main biological pathways that were consistently enriched in SLE patients. Subsequently, a meta-signature consisting of 37 genes involved in diverse processes was generated. Each gene that was selected met the original criteria, was involved in at least one relevant pathway and had a fold change of over 2 in at least one of the datasets. This signature differentiated well between children with SLE and healthy controls in a fifth independent dataset.

Comparison of SLE with healthy controls could generate insights into the underlying immune dysfunction and thus help identify therapeutic targets for SLE. Signatures that were found to be consistently enriched between the different datasets included interferon signaling, corroborating and extending earlier findings, and interleukin-10 signaling, which may reflect a dysregulated inflammatory process linked to humoral immune activation and Janus kinase (JAK)/signal transducer and activator of transcription (STAT) signaling. Finally, a glucocorticoid receptor signaling signature was also implicated, although corticosteroid therapy, frequently used for SLE, may be a confounding factor. Specific signatures such as the (low density) neutrophil, immunoglobulin and lymphopenic signatures that were previously reported to be part of one of the datasets [5] were not listed.

Interestingly, no differences between the children and adults with SLE were observed [1]. This finding supports observations from studies directly and indirectly comparing cohorts of children and adults with SLE, which revealed that no known unique physiological or genetic pathways were identified that can explain the variability in the disease phenotypes [6].

Overall, the pathway-based approach of Arasappan and colleagues [1] seems to offer a simple and valuable way to increase the power of microarray data meta-analysis. Using different pathway-level stringencies and approaches, such as Gene Set Enrichment Analysis, PANTHER and Metacore for confirmation of pathway signatures, may increase the robustness of the results. The value of the different meta-analysis methods will become apparent in comparative studies. Such benchmarking, together with incorporation of properly annotated demographic and clinical data, would allow optimization of meta-analysis approaches.

Next steps

The future challenge is to use meta-analysis strategies to identify gene signatures and biomarkers that can improve measures for diagnosis and disease activity in SLE.

In an attempt to use a blood leukocyte gene expression profile to improve the diagnosis of SLE, Chaussabel and colleagues [7] went one step further and compared a dataset from SLE with datasets from several other diseases, including those that show an interferon signature. To compare datasets across multiple diseases, they used a custom meta-analysis strategy for diagnostic biomarker selection using statistical significance ($P <$ 0.01), rather than the preferable gene expression size effects between each group of patients versus healthy controls. Subsequent selection of genes that reached significance for the comparison between SLE patients and healthy controls, and not compared to healthy subjects in the other diseases, led to the identification of an SLE-specific 'diagnostic signature' that differentiates SLE patients from patients with diseases that also show an interferon signature [7].

The SLE disease activity index (SLEDAI) consists of a series of measures that is sometimes difficult to obtain, so development of a simple and objective index based on a blood leukocyte gene expression profile would be useful. Arasappan and colleagues [1] provide evidence that their 37-gene meta-signature discriminates between patients with a low (<3) and high (>3) SLEDAI score, confirming results from other independent studies. One such study was performed by Chaussabel and colleagues [8], who used a 'modular analysis framework' based on the identification of coordinately expressed genes in disease datasets that form transcriptional modules. Their approach predicted severity of disease more accurately than the SLEDAI in some cases, demonstrating that the blood leukocyte gene expression profile might be useful for discovering diagnostic and prognostic biomarkers and monitoring disease progression.

Towards a systems biology approach

Unique to transcriptome analyses is the identification of gene signatures that represent biological networks, such as the interferon system, that are relevant in disease

pathogenesis and thus provide a starting point for a systems biology approach. Successive research activities on these networks, together with approaches using complementary platforms such as (epi)genetics, multiplex fluorescence-activated cell sorting and advanced metabolomics/proteomics, will provide a complete insight into the mechanism and other network components of processes and pathways relevant to disease. For example, interferon-based genetic studies led to the identification of polymorphisms that are strongly associated with SLE in genes encoding interferon regulatory factor 5 (*IRF5*), STAT4, interleukin-1 receptor-associated kinase 1 (*IRAK1*), autophagy protein 5 (*ATG5*) and three prime repair exonuclease 1 (*TREX1*), genes that are all connected with deregulated interferon activity [9]. Proteomics on downstream components revealed that a composite chemokine score for the interferon-regulated chemokines CXCL10 (IP-10), CCL2 (MCP-1) and CCL19 (MIP-3B) in patients with a SLEDAI of 4 or less were predictive of a lupus flare over the ensuing year [10].

Thus, besides identifying clinically relevant transcriptome markers, DNA-microarray technology provides a basis for an evidence-based systems biology approach to delineate pathogenic processes and reveal other relevant markers. Meta-analysis methods will be instrumental in helping to select those exploratory markers for further biomarker validation, which will pave the way for clinical development and benefit patients.

Abbreviations
SLE, systemic lupus erythematosus; SLEDAI, SLE disease activity index.

Competing interests
The authors declare that they have no competing interests.

References
1. Arasappan D, Tong W, Mummaneni P, Fang H, Amur S: **Meta-analysis of microarray data using a pathway-based approach identifies a 37-gene expression signature for systemic lupus erythematosus (SLE) in human peripheral blood mononuclear cells.** *BMC Med* 2011, **9:**65.
2. Brazma A: **Minimum Information About a Microarray Experiment (MIAME)-- successes, failures, challenges.** *ScientificWorldJournal* 2009, **9:**420-423.
3. MAQC Consortium: **The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models.** *Nat Biotechnol* 2010, **28:**827-838.
4. Shi L, Tong W, Fang H, Scherf U, Han J, Puri RK, Frueh FW, Goodsaid FM, Guo L, Su Z, Han T, Fuscoe JC, Xu ZA, Patterson TA, Hong H, Xie Q, Perkins RG, Chen JJ, Casciano DA: **Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential.** *BMC Bioinformatics* 2005, **6 Suppl 2:**S12.
5. Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J, Pascual V: **Interferon and granulopoiesis signatures in systemic lupus erythematosus blood.** *J Exp Med* 2003, **197:**711-723.
6. Mina R, Brunner HI: **Pediatric lupus--are there differences in presentation, genetics, response to therapy, and damage accrual compared with adult lupus?** *Rheum Dis Clin North Am* 2010, **36:**53-80.
7. Chaussabel D, Allman W, Mejias A, Chung W, Bennett L, Ramilo O, Pascual V, Palucka AK, Banchereau J: **Analysis of significance patterns identifies ubiquitous and disease-specific gene-expression signatures in patient peripheral blood leukocytes.** *Ann N Y Acad Sci* 2005, **1062:**146-154.
8. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, Stichweh D, Blankenship D, Li L, Munagala I, Bennett L, Allantaz F, Mejias A, Ardura M, Kaizer E, Monnet L, Allman W, Randall H, Johnson D, Lanier A, Punaro M, Wittkowski KM, White P, Fay J, Klintmalm G, Ramilo O, Palucka AK, Banchereau J, Pascual V: **A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus.** *Immunity* 2008, **29:**150-164.
9. Pascual V, Chaussabel D, Banchereau J: **A genomic approach to human autoimmune diseases.** *Annu Rev Immunol* 2010, **28:**535-571.
10. Bauer JW, Petri M, Batliwalla FM, Koeuth T, Wilson J, Slattery C, Panoskaltsis-Mortari A, Gregersen PK, Behrens TW, Baechler EC: **Interferon-regulated chemokines as biomarkers of systemic lupus erythematosus disease activity: a validation study.** *Arthritis Rheum* 2009, **60:**3098-3107.