A Transcriptome Fingerprinting Assay for Clinical Immune Monitoring

Matthew C Altman[1,2,*], Nicole Baldwin[3], Elizabeth Whalen[1] , Taha Al-Shaikhly[2], Scott Presnell[1], Prasong Khaenam[1], Vivian H Gersuk[1], Laurent Chiche[4], Noemie Jourde-Chiche[5], J Theodore Phillips[3], Goran Klintmalm[3], Anne O'Garra[6,7], Matthew Berry[8], Chloe Bloom[6], Robert J Wilkinson[9,10,11], Christine M Graham[6], Marc Lipman[12], Ganjana Lertmemongkolchai[13], Farrah Kheradmand[14], Asuncion Mejias[15], Octavio Ramilo[15], Karolina Palucka[3,16], Virginia Pascual[3,17], Jacques Banchereau[3,16], Damien Chaussabel[1,18*]

1 *Systems Immunology, Benaroya Research Institute, Seattle, Washington, USA*

2 *Division of Allergy and Infectious Diseases, University of Washington, Seattle, Washington, USA*

3 *Baylor Institute for Immunology Research and Baylor Research Institute, Dallas, Texas, USA*

4 *Department of Internal Medicine, Hospital Europeen, Marseille, France*

5 *Aix-Marseille University, C2VN, INSERM 1263, INRA 1260, Marseille, France*

6 *Laboratory of Immunoregulation and Infection, The Francis Crick Institute, London, UK*

7 *National Heart and Lung Institute, Imperial College London, London W2 1PG, UK*

8 *Respiratory Medicine, Imperial College Healthcare NHS Trust, London, UK*

9 *The Francis Crick Institute, London, UK*

10 *Department of Medicine, Imperial College, London, UK*

11 *Wellcome Center for Infectious Diseases Research in Africa and Department of Medicine, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town Observatory 7925, Republic of South Africa*

12 *UCL Respiratory, Division of Medicine, University College London, London, UK*

13 *Centre for Research and Development of Medical Diagnostic Laboratories, Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, Thailand*

14 *Baylor College of Medicine, Houston, Texas, USA*

15 *Nationwide Children's Hospital and the Ohio State University School of Medicine, Division of Pediatric Infectious Diseases, Columbus, Ohio, USA*

16 *The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA*

17 *Weill Cornell Medicine, New York, New York, USA*

18 *Systems Biology, Sidra Medicine, Doha, Qatar*

[*]To whom correspondence may be addressed:

Matthew C Altman, MD, Systems Immunology Division**,** Benaroya Research Institute, 1201 Ninth Avenue, Seattle, WA 98101, USA. Tel. +1 206 287 5648, Fax. 206 287 5682, E-mail: maltman@benaroyaresearch.org

Damien Chaussabel, PhD, Systems Biology Department**,** Sidra Medical and Research Center, PO Box 26999 Al Luqta Street, Doha, Qatar. Tel. +974 4003 7395, E-mail: dchaussabel@sidra.org

Keywords: human immunology; transcriptome; gene expression; systems biology; systemic lupus erythematosus; pregnancy

ABSTRACT

Background: While our understanding of the role that the immune system plays in health and disease is growing at a rapid pace, available clinical tools to capture this complexity are lagging. We previously described the construction of a third-generation modular transcriptional repertoire derived from genome-wide transcriptional profiling of blood of 985 subjects across 16 diverse immunologic conditions, which comprises 382 distinct modules.

Results: Here we describe the use of this modular repertoire framework for the development of a targeted transcriptome fingerprinting assay (TFA). The first step consisted in down-selection of the number of modules to 32, on the basis of similarities in changes in transcript abundance and functional interpretation. Next down-selection took place at the level of each of the 32 modules, with each one of them being represented by four transcripts in the final 128 gene panel. The assay was implemented on both the Fluidigm high throughput microfluidics PCR platform and the Nanostring platform, with the list of assays target probes being provided for both. Finally, we provide evidence of the versatility of this assay to assess numerous immune functions *in vivo* by demonstrating applications in the context of disease activity assessment in systemic lupus erythematosus and longitudinal immune monitoring during pregnancy.

Conclusions: This work demonstrates the utility of data-driven network analysis applied to large-scale transcriptional profiling to identify key markers of immune responses, which can be downscaled to a rapid, inexpensive, and highly versatile assay of global immune function applicable to diverse investigations of immunopathogenesis and biomarker discovery.

BACKGROUND:

Traditionally the immune system has been viewed as playing a beneficial role in control of infection and a detrimental role in autoimmunity and allergic processes. More recently it has been appreciated to also have critical functions in a far wider range of common diseases including obesity, atherosclerosis, dementia, and numerous cancers among others[1-4]. The immune system acts as a highly interconnected network of cellular and humoral interactions and it is through either appropriate function or malfunction of network components that the immune system underpins these diverse human diseases [5-9].

A better understanding of individual immune processes has facilitated numerous interventions that can alter immune responses and ameliorate outcomes. However cost-effective and standardized tools available to practitioners and clinical trialists that capture the complexity of immune responses and that could be used for monitoring the immune status of patients are lacking. Global monitoring of the immune system, even in clinical trials of immunotherapies, is often missing due to the complexity of implementing systems approaches.

Immune responses are highly complex, dissecting out the reproducible global patterns of the immune system is critical to developing improved methods of immune monitoring [10]. In this regard, systems-scale analyses of global architecture in both normal and pathologic immune function is necessary. Different technologies under development and used in research have numerous strengths but also distinct challenges, including cost of reagents or instruments, complexity of the assay workflow, and complexity of data analysis and interpretation.

We previously described the construction of a third-generation modular repertoire, compromised of 382 modules, and is representative of 16 immune states. We also described transcriptome fingerprinting as a novel and useful visualization scheme [11]. Here we take our work a step further, and describe the development of a cost-effective and practical, targeted transcriptome fingerprinting assay (TFA) (**Figure 1**), while preserving its capability of monitoring the same modular repertoire of immune responses. This assay is by design meant as a generic assay suitable for immune profiling across multiple states of health and disease. We demonstrate the utility of creating such an assay through a purely data-driven network analysis approach to identify core functional immune pathways. Our results show the complex molecular interactions in immune pathogenesis but also reveal a redundancy of core immune circuits.

This assay highly simplifies the analysis and interpretation of gene transcription data in immunologic diseases. Several applications are envisioned, including assessment of disease activity, health monitoring, pre-symptomatic detection of disease, and biomarker discovery. In this manuscript we show application of the assay to assess disease activity in systemic lupus erythematosus (SLE) and for longitudinal immune monitoring during pregnancy. This assay is released as an “open resource”, including a complete list of reagents, detailed procedures, and source code for data analysis.

RESULTS:

In our third-generation modular repertoire, the 382 modules showed an inherent variability. We sought to show that the variability apparent in the modules can be reduced by using a subset of the module genes, demonstrating that there are core genes within each module that best reflect perturbations of that pathway. It also produces sets of genes that support rapid and cost effective transcriptional profiling. In order to find a representative subset of modules that best reflects the variability seen across the source data, we partitioned the modules with Hartigan's K-means algorithm using the jump statistic [12] to determine an appropriate number of clusters and to reduce granularity, resulting in 38 subgroups. The module closest to the mean vector in each subgroup was selected to represent that subgroup, and if a subgroup did not contain at least one module in one of the 16 diseases showing at least 25% of genes up- or down-regulated, it was excluded from selection. This left 32 modules representative of the 382 original modules (**Figure 2**). From each of the 32 modules we then selected 4 representative probes by ranking all probes according to the distance of each probe from the module's mean probe vector. The highest ranking probes that had gene symbols unique to a module were selected.

This unbiased process selected out modules most representative of each cluster and genes most representative of each module, resulting in a subset of 32 modules and 4 genes per module for a total of 128 genes representative of the diversity among the 382 modules and 14,502 total transcripts profiled. These 32 modules/128 genes were chosen to represent the "transcriptome fingerprint" used for assessment and monitoring of immunologic disease states (Table 1). This scaling down process allows the implementation of targeted assays, which, given the markedly reduced volumes of data generated per experiment, are cost-effective and allow higher sample throughput, faster turnaround and streamlined analysis and interpretation, while maintaining profiling of a broad repertoire of immune gene signatures.

Identification of biomarkers for disease activity in SLE

We first sought to show that a transcriptome fingerprinting assay (TFA) panel when measured by high-throughput qPCR, shows high correlation with gene expression data derived from microarray. For this we used a cohort of adult patients with SLE, which was independent from the cohort used for module construction [13]. Whole blood samples from 24 SLE patients with varying stages of disease activity as well as 15 healthy age and gender matched controls were used for analysis. Samples were selected that had been drawn prior to initiation of new immunosuppressive treatment to reflect disease activity before therapy. Whole blood derived RNA samples taken at an initial visit were used to generate both microarray and qPCR data. High correlation values were observed between microarray and qPCR data for all genes and modules that had FC values greater than 1.5 for the ratio between SLE and healthy control expression levels (**Figure 3a**).

Next, differences in immunologic pathways potentially useful for stratifying the SLE patients were evaluated. Four modules that are biologically relevant to SLE showed distinct up-regulation in SLE relative to healthy controls. These were M12.15, annotated as a cell cycle module, M8.3 (type 1 interferon), M15.86 (interferon), and M10.4 (neutrophil activation). Module M12.15 (cell cycle) showed a high degree of correlation with clinical disease activity as measured by the SLE Disease Activity Index (SLEDAI) score (R=0.793, P=6.03E-6) [14]. This module showed a higher level of correlation than traditional markers of disease activity – anti-double-stranded DNA titers (R=0.636, P=0.001) and C-reactive protein (CRP) (R=0.327, not significant) – which had been measured at the same time (**Figure 3b**) [15]. The four representative genes within module M12.15 are linked to SLE. TYMS is an enzyme critical for folate metabolism and polymorphisms in this and related genes have previously been associated with SLE [16]. CD38 has been shown to be expressed in higher levels on circulating lymphocytes in active SLE [17]. TNFRSF17 (also called BCMA) is a B-cell maturation antigen that has been shown important for B-cell development and SLE pathogenesis in mouse models [18]. MZB1 (previously MGC29506) has been shown to be involved in immunoglobulin heavy chain biosynthesis [19]. Our finding demonstrates that four genes representative of a signature important to B cell development show very high correlation with SLE disease activity and that TFA could present potential clinical utility for assessment of disease activity.

Abundance of transcripts representative of two interferon modules, M8.3 and M15.86, was also increased in the SLE samples relative to control. Several groups have previously reported an increase in type I interferon (IFN) regulated genes in a subset of SLE patients (reviewed in [20] and [21]). More recently, whole-genome transcriptional profiling has demonstrated a higher degree of complexity in interferon activity with an apparent gradient of interferon responses in SLE [13, 22]. Here our findings support this observation; since patients with SLE included in our study could be similarly stratified based on either absence of interferon signature, activation of a single interferon module M8.3, or activation of both interferon modules M8.3 and M15.86, in an apparent sequential pattern (**Figure 3c**). The genes representative of M8.3, ISG15, IFI44, LY6E, and XAF1, are all well characterized as IFN-alpha inducible genes. The genes representative of M15.86, MOV10, TIMM10, KIAA1618, and GALM are less well characterized and may represent either a related or distinct interferon response seen in a subset of SLE patients in parallel with a relatively saturated IFN-alpha response. TFA provides a straightforward framework to further investigate this differential interferon response in SLE.

Module M10.4 is composed of neutrophil specific genes, but did not show significant correlation with participants' absolute neutrophil counts collected at the same time (R=0.33, P=0.144), suggesting this module reflects alterations in neutrophil function and activity more than quantity. The four representative genes that were selected out of the fifteen genes constituting M10.4 encode well characterized proteins important to neutrophil function; CEACAM6 and

CEACAM8 are cell-adhesion proteins on neutrophils essential for neutrophil adhesion and migration [23]; DEFA4 is a defensin peptide, and ELANE or neutrophil elastase is a serine protease, both are contained in neutrophil azurophil granules. Many, but not all patients show significant increase in abundance of transcripts representative of M10.4, suggesting abnormal neutrophil function in only a specific subset of untreated SLE (**Figure 3d**). There is growing evidence for alterations in neutrophil chemotaxis, phagocytosis, superoxide production, and apoptosis in subgroups of SLE patients and investigation of neutrophil activity is an active area of research in SLE pathogenesis[24-26]. Our assay agrees with findings that neutrophil dysregulation may be observed in only a specific subset of SLE and suggests that this gene set could serve as a marker to identify SLE patients with neutrophil dysregulation.

Longitudinal monitoring of immune status in pregnancy

To determine the baseline expression of TFA modules in healthy adults and to assess their utility for monitoring of immunologic changes, the assay was run on samples collected longitudinally from 18 healthy non-pregnant volunteers and 12 healthy pregnant women. Samples were collected at 2-week intervals for up to 28 weeks. In the pregnant women, sample collection started at ~10 weeks into the pregnancy. All of these women had uncomplicated term deliveries of healthy infants. We monitored changes in gene expression of these 32 modules over time to assess for consistent changes attributable to pregnancy which is likely to mediate progressive immunological and physiological changes over time.

Investigation of the healthy controls demonstrated no significant differences in expression levels of these 128 genes according to gender, age, or time point (**Supplemental Figure 1**). Therefore, these 18 individuals were used to define reference ranges for expression of these 128 genes. To support comparisons amongst modules, all expression values were scaled to a mean of 1, and confidence intervals defined by the standard errors of the healthy controls. Most modules (26/32) had narrow confidence intervals of +/- 10% or less, demonstrating these genes show relatively modest fluctuation in healthy individuals (Table 2). Only one module, M8.3 (type 1 interferon), showed particularly high variability (+/- 32%), which we believe could be related to viral infections in some of the controls throughout the course of monitoring or else may reflect higher normal temporal variability of expression of these genes.

Samples from the pregnant women were tested for consistent changes during the course of monitoring as well as group differences compared to healthy controls. After multiple testing correction, 7 modules, M10.4, M8.2, M12.11, M14.53, M13.16, M12.2, and M15.55, showed a significant linear change over time. Another 4 modules, M14.76, M15.102, M12.4, and M12.3, were significantly different from controls over the period of time monitored suggesting a change in expression occurred in the first 10 weeks of pregnancy prior to the start of sample collection (**Figure 4** and **Supplemental Figure 3**).

Several of these modules represent immune functions that are known to change during healthy pregnancy. M8.2 and M15.102 are both relevant to prostanoid metabolism. Prostanoids are critical for cervical and uterine development in preparation for delivery [27]. M8.2 shows a linear increase in expression throughout pregnancy with fairly narrow confidence intervals (+/- 13%) and based on the genes in M8.2 likely relates to increasing prostanoid production. M15.102 is decreased relative to controls throughout monitoring. M15.102 includes both GPR44 (also called PTGDR2) a prostaglandin D2 receptor and IL5RA a subunit of the IL5 receptor, both of which play critical roles in T helper cell mediated immune responses, as well as OLIG2 and PRSS33, which are less well characterized. The decrease in this module may reflect a change in lymphocyte function in relation to changing levels of prostanoids.

M10.4 is a neutrophil activation module. It increases dramatically during pregnancy to more than 4-fold on average. This is likely in part due to the known increase in neutrophil numbers in the peripheral blood during the 2nd and 3rd trimesters, but likely also reflects changes in neutrophil function as previously observed [28] since the genes composing this module are specific to neutrophil activation as discussed earlier. Module M12.2 (composed of genes ALDH2, CEBPA, EMILIN2, and KYNU) is a monocyte module. Monocytes generally do not change in number during pregnancy but show increased activation in the circulation [29]. The 7 other modules that show different expression patterns compared to healthy controls represent several other biological processes including novel findings that compel further investigation (**Supplemental Figure 2**). Taken together our findings show that the TFA assay results are stable over time in non-pregnant healthy adults and can detect progressive immunologic changes during the course of a healthy pregnancy. This provides a baseline for further investigation of immunologic changes that can occur during both healthy and complicated pregnancies.

DISCUSSION:

We present here, the design and implementation of modular transcriptional repertoire-based targeted assays. It is based on our prior work in constructing a third-generation modular repertoire in clinical immunology, whereby variation in abundance of blood RNA was captured through the construction of co-clustered transcriptional modules. The modular repertoire, was representative of 16 immune states (16 instead of 7 and 8 in earlier generations) [30, 31], and used as an input transcriptome profiles of nearly 1000 subjects. That approach identified 382 modules showing co-clustering across a wide range of immune conditions, while others appeared to be more condition-specific. In this work, we show that modules can be reduced to representative genes in a purely data-driven fashion that does not depend on a priori knowledge about the genes or clinical states.

Using such approach, we found that useful representative genes of a functional pathway may not be canonical genes, and that gene selection through a data-driven network analysis approach is powerful for novel discovery and assay development. This method, which we have called a "transcriptome fingerprinting assay" or TFA, enabled down-scaling from complicated genome-wide expression profiling to rapid and cost-effective qPCR and molecular barcoding platforms. Proof of principle was provided for disease pathogenesis, biomarker discovery, and longitudinal monitoring applications. TFA was employed to investigate immune perturbation in SLE and pregnancy. We were first able to establish the high degree of correlation between TFA and microarray data, and to demonstrate stability of the TFA gene signature in healthy adults over time. More importantly we were also able to demonstrate the ability of this assay to detect both known and novel biological changes. In the case of SLE, confirming for instance the differential expression of interferon genes, and adding evidence regarding neutrophil dysregulation in a subset of patients. We also found that a cell cycle module shows a very high degree of correlation with SLE disease activity, which warrants further investigation as a potential disease biomarker. Clinical utility may be found through combination of these modules to provide rapid and effective means to stratify and monitor SLE patients.

In the setting of pregnancy, numerous modules involved in prostanoid metabolism, neutrophil activation, and monocyte activation were found to change in a coherent fashion throughout the course of the second and third trimester. Some of those modules constitute a means to quantify immune changes that are known to take place during pregnancy while others appear to track changes that have not previously been recognized and that will need to be further characterized. Indeed, currently there are no biomarkers for two of the most common adverse pregnancy outcomes: preterm labor/delivery and preeclampsia/eclampsia [32, 33]. Risk factors for preterm labor and delivery are believed to act through multiple immune pathways including altering eicosanoid metabolism and increasing prostaglandin production [34, 35] and through changes in neutrophil cytokine production [36]. Similarly changes in eicosanoid metabolism play a major role in preeclampsia and eclampsia [37], and both neutrophil and monocyte activation are thought to be important in the pathophysiology of this condition [28, 38, 39]. We hypothesize that longitudinal measurement using TFA modules, in particular those we have demonstrated to have coherent change throughout pregnancy, could be used for a better understanding and eventually early detection of maternal and perinatal complications such as preterm birth and pre-eclampsia. Testing of this hypothesis is currently underway.

CONCLUSION:

To conclude, this work demonstrates the utility of purely data-driven network analysis applied to large-scale transcriptional profiling datasets to identify key markers of immune responses. From this approach we have developed a transcriptome fingerprint of the immune system based on a non-systems scale assay, which is applicable to investigation of immunopathogenesis, longitudinal monitoring, and biomarker discovery. Sample acquisition for this assay is straightforward as blood can be collected by venipuncture or finger stick [40] and requires no onsite processing. The TFA assay is cost-effective, generates a manageable volume of data, and does not require sophisticated bioinformatics infrastructure and pipelines for analysis.

Notably the successful use of both a PCR based assay and a well-established molecular barcoding technology (NanoString) confers additional advantages, since both technologies are known for sensitivity, robustness, and ease of use. Furthermore PCR is widely used in clinical diagnostic and research settings, which would allow our assay to be easily adopted. We are publishing this assay as an "open resource", including a complete list of reagents and source code for data analysis. This should facilitate third party implementation of the assay and hopefully encourage re-sharing of iterative improvements of its design and of the downstream analytic pipeline. Indeed, taken together, the development of streamlined "Omics-based" assays should contribute to a wider adoption of systems approaches, or in this case "systems-based" approaches.

METHODS

Modules/Genes Selection - Downscaling to transcriptome fingerprinting

The 382 modules comprising our third generation modular repertoire were grouped using Hartigan's K-means algorithm and using the jump statistic to determine an appropriate number of clusters [12], resulting in 38 subgroups. The module closest to the mean vector in each subgroup was selected to represent that subgroup. If a subgroup did not contain at least one module in one of the 16 diseases showing at least 25% of genes up- or down-regulated, it was excluded from selection. This left 32 modules representative of the 382 original modules. From each of those 32 modules, 4 representative probes were selected by ranking all probes according to the distance of each probe from the module's mean probe vector and the number of presence calls per sample group (detection $P< 0.01$). The highest-ranking probes that had gene symbols unique to a module were selected.

Microarray data generation (SLE cohort)

Globin mRNA was depleted using the GLOBINclear™ (Thermo Fisher Scientific). Globin-reduced RNA was amplified and labeled using the Illumina TotalPrep-96 RNA Amplification Kit (Thermo Fisher Scientific). Biotin-labeled cRNA was hybridized overnight to Human HT-12 V4 BeadChip arrays (IIlumina), which contains >47,000 probes, and scanned on an Illumina BeadStation to generate signal intensity values.

TFA data generation

For the SLE cohort, a quantitative reverse transcription PCR platform was used. Globin reduced RNA was reverse-transcribed to cDNA using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific), followed by specific target preamplification for 14 cycles in the presence of a pool of 136 primer pairs, including 8 reference genes (Supplemental Methods) (DELTAgene Assays, Fluidigm). Preamplified cDNAs were treated with Exonuclease I (New England Biolabs) to remove unincorporated primers and the preamplified cDNAs and detection assays were loaded onto a 96.96 Dynamic Array IFC (Fluidigm). Real-time PCR was run using EvaGreen dye (Bio-Rad) for detection on a BioMark HD System (Fluidigm). Analysis was performed using the Real-Time Analysis Software package (Fluidigm) to determine cycle threshold (Ct) values, using linear (derivative) baseline correction and auto-detected, assay-specific threshold determination.

For the pregnancy cohort, a NanoString assay was used. 100ng of total RNA was hybridized overnight (18 h) to target genes contained in a custom gene expression nCounter Plex2 for GEx NanoString Assay (Supplemental Methods), following the manufacturer's Gene Expression Assay protocol. Enrichment of hybridized reporter/capture complexes and RNA target was carried out using SamplePrep Station and signal detection was carried out in an nCounter Digital Analyzer set for high-resolution scanning. NanoString data analysis guidelines were followed to carry out normalization to assay positive controls and to subtract background noise. Normalization to housekeeping genes included in custom gene panel (Supplemental Methods) was carried out using housekeeping-gene global geometric mean approach. Resulting normalized values were reported for downstream statistical analysis.

Statistical Analyses

Two-group comparisons (t-tests) were run on log2 FC values between SLE and healthy controls to determine modules that showed significant differences between the two groups. For the longitudinal pregnancy data, mixed effects models, using the lme4 package in R [41], were run to compare pregnancy versus healthy controls over time. Principal component analysis (PCA) was performed on all healthy controls but no significant differences were found according to gender, age, or time of sample collection, so all healthy controls samples were included.

Abbreviations

BIIR: Baylor Institute for Immunology Research

COPD: Chronic Obstructive Pulmonary Disease

CRP: C-Reactive Protein

Ct: Cycle threshold

FC: Fold Change

FDR: False Discovery Rate

GO_BP: Gene Ontology Biologic Process

GO_MF: Gene Ontology Molecular Function

HIV: Human Immunodeficiency Virus

IFN: Interferon

IRB: Institutional Review Board

KEGG: Kyoto Encyclopedia of Genes and Genomes

MS: Multiple Sclerosis

PCA: Principle Component Analysis

PCR: Polymerase Chain Reaction

PID: Primary Immune Deficiency

RF: Random Forest

ROC: Receiver Operating Characteristic

RSV: Respiratory Syncytial Virus

SLE: Systemic Lupus Erythematosus

SLEDAI: SLE Disease Activity Index

SOJIA: Systemic onset Juvenile Idiopathic Arthritis

TB: Tuberculosis

TBD: To Be Determined

TFA: Transcriptome Fingerprinting Assay

Declarations

Ethics approval and consent to participate

Each of the studies contributing samples for this manuscript was independently approved by the BIIR IRB (IRB #'s 009-240, 006-177, 002-197, 009-257, H-18029, HE-470506, 011-173.

Availability of data and material

Raw gene expression data was deposited in the Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/, under the accession GSE100150.

Competing interests

The authors declare no conflicts of interest.

Funding

This project has been funded with Federal funds from the National Institutes of Health under contract number U01AI082110.

Authors' contributions

Microarray data were generated at the Baylor Institute for Immunology Research and Benaroya Research Institute. qPCR data were generated at the Benaroya Research Institute. N.B. developed the modular framework. M.C.A, N.B, P.K, V.H.G, and D.C developed the TFA panel. M.C.A, P.K, and E.W analyzed data. L.C, N.J, J.T.P, and R.A.L contributed the samples run on TFA. J.T.P, G.K, M.F.L, H.B.R, A.O, M.B, C.B, M.L, R.W, C.G, G.L, M.E.C, J.S, F.K, K.P, A.M, O.R, V.P, and D.C contributed the microarray data for module generation. P.S.L, R.A.L, J.B, C.Q, S.P, and D.C.A provided support and contributed to study design. M.C.A and D.C wrote the manuscript.

Acknowledgements

We thank Dr. G Nepom for helpful comments on the experimental design. We thank Quynh-Anh Nguyen, Kimberly O'Brien, Dimitry Popov, Michael Mason, and Cate Speake for technical assistance. We thank Brad Zeitner for development of software.

FIGURES:

Figure 1: Overview of the module repertoire construction and targeted assay development approach. All details are provided in the main text and supplemental methods sections. Briefly, blood transcriptional module repertoire construction takes a collection of transcriptome datasets as input. In this case 16 datasets constituted by 985 individual transcriptome profiles spanning a wide range of immunological and physiological states. Clustering behavior of gene pairs is recorded for each independent datasets and the information complied in a co-clustering table. Subsequently the co-clustering table serves as input for the generation of a co-clustering graph, where nodes are the genes and edges represent co-clustering events. Next the largest, most densely connected subnetworks among a large network constituted of 15,132 nodes are identified mathematically and assigned a module ID. The genes constituting this module are removed from the selection pool and the process is repeated. The resulting framework of 382 modules served as a basis for the development of targeted assay. This involves two major steps. First, the selection of representative modules among the 382 modules constituting the framework. Second, the selection of representative probes among those modules. The process can be adjusted according to practical constraints, such as assay throughput and cost. In our case the selection of 32 modules out of the original set of 382, and of 4 representative genes from each of the 32 modules yielded a 128-gene fingerprinting assay.

Figure 2: Patterns of blood transcript abundance observed across 16 disease or physiological states (A) Depicted is the expression pattern of the gene members in each of the 382 modules (columns) across all 16 disease states (rows). Each pixel represents the percent of probes within that module that show a significant difference in expression between the disease group and the control group within that microarray dataset. The color scale ranges from 100% up-regulated (full red) to 100% down-regulated (full blue). The modules are clustered into 38 distinct subgroups separated by black vertical lines according to similarity of expression pattern across the 16 disease states.

(B) The expression pattern of the gene members in each of the 32 representative modules of the 382 original modules/38 clusters depicted in **(A)** representing the downscaled 32 modular repertoire.

Figure 3: TFA analysis of SLE demonstrates activation of cell cycle, interferon, and neutrophil pathways.

(A) Each point represents the fold change of a single gene comparing the average of SLE patients to the average of Controls as measured by qPCR (x-axis) and microarray (y-axis). Average fold change values were similar and proportional across technologies. The color of each point represents the Pearson’s correlation of fold change values among SLE patients for a given gene as measured by qPCR and by microarray. Those genes with high average fold change values also showed very high levels of correlation across platforms.

(B) Fold change value of module M12.15 (average of 4 genes) plotted versus the SLEDAI score for each SLE patient. Correlations and *p* values were calculated using Spearman's rank correlation coefficient.

(C) Interferon module fold change values are shown for each SLE patient (n=24) compared to the average of the healthy controls. Each point represents the module fold change of a single SLE patient. Patients are ordered according to increasing FC of M8.3. The greater the intensity of each point, the greater the fold change. White FC≤2. Full red ≥20.

(D) Neutrophil module fold change values are shown for each SLE patient (n=24) compared to the average of the healthy controls. Each point represents the module fold change of a single SLE patient. Patients are ordered according to increasing FC of M10.4. The greater the intensity of each point, the greater the fold change. White FC≤2. Full red ≥20.

Figure 4: Changes in blood transcript abundance measured during the course of uncomplicated pregnancies. Average FC expression levels referenced to healthy controls from the blood of 12 healthy pregnant women and 18 healthy non-pregnant controls. Shown are 4 of the 11 significant modules (see **Supplemental Figure 3** for the other 7). A linear mixed effects model was fit to the longitudinal data from pregnant women and healthy controls to determine if there was a trend over time for the pregnant women (time *p*-value) and if there was a significant difference between pregnant women and healthy controls over time (condition *p*-value).

Supplemental Figure 1: TFA gene expression in healthy controls.

(A) Gene expression values from all healthy controls were used for principal component analysis. Scores from principal component 1 (PC1) and 2 (PC2) were plotted with these 2 components explaining about 19% of the variability in the data. There were no group differences attributable to gender among the expression values (red=female, blue=male).

(B) Plot of PC1 and PC2 from principal component analysis of gene expression values for all female controls, which again explains approximately 19% of the variability in the data. There were no group differences attributable to age (as a surrogate for child bearing status) in the expression values (red=women <40yo, black=women>40yo).

(C) No significant differences were found between healthy control samples obtained in different seasons as shown in this principal component plot (blue=winter, yellow=spring, green=summer, and orange=fall).

Supplemental Figure 2: TFA assessment during healthy pregnancy demonstrates longitudinal immunological changes.

Average FC expression levels referenced to healthy controls from the blood of 12 healthy pregnant women and 18 healthy non-pregnant controls. Shown are 7 of the 11 significant modules (see Figure 4 in the main text for the other 4). A linear mixed effects model was fit to the longitudinal data from pregnant women and healthy controls to determine if there was a trend over time for the pregnant women (shown as the time p-value) and if there was a significant difference between pregnant women and healthy controls over time (shown as the condition p-value).

TABLES:

Table 1: TFA modules

ID	Module title	Number of unique genes	TFA genes	Module Variability in Healthy Controls	Disease Activity Up >50%	Disease Activity Down >50%
M8.2	Prostanoids	36	CTDSPL, SH3BGRL2, TSPAN33, TSPAN9	14.1%	Staph, RSV	MS, HIV
M8.3	Type 1 Interferon	17	ISG15, IFI44, XAF1, LY6E	32.2%	HIV, SOJIA, RSV, Influenza, SLE, TB, Transplant	
M10.2	Protein synthesis	19	HBB, LAIR1, OAZ1, RPS12	3.8%		MS
M10.4	Neutrophil activation	13	CEACAM8, DEFA4, CEACAM6, ELANE	15.6%	HIV, Staph, Burkholderia, RSV, Transplant, Kawasaki, Pregnancy	
M12.2	Monocytes	44	ALDH2, CEBPA, EMILIN2, KYNU	7.7%	MS, Staph, COPD	
M12.3	Cell cycle	70	ELP3, LANCL1, NUP160, TTC4	5.3%	MS	B-Cell Deficiency, Pregnancy, Transplant, Staph, Burkholderia
M12.4	Gene Transcription	62	CCDC12, C19ORF53, E4F1, NDUFA8	3.0%	MS, HIV	SOJIA, Pregnancy, Burkholderia
M12.5	Protein modification	91	INTS10, CCDC16, RPS6KB1, ZFYVE20	3.0%		Transplant, Staph, Burkholderia
M12.11	Erythrocytes	24	RAD23A, TRAK2, SIAH2, RPIA	8.6%	Staph, SOJIA, RSV, Transplant, Melanoma	
M12.15	Cell cycle	17	CD38, MGC29506, TNFRSF17, TYMS	7.2%	HIV, SLE, Kawasaki	
M13.16	Cytokines/chemokines	39	KCNJ2, ALPK1, GK, LRG1	7.9%	Staph, Burkholderia, SOJIA, Influenza, TB	
M13.19	TBD	63	MAP3K5, CEP350, PIK3CG, PPTC7	3.5%	SOJIA	RSV, Transplant
M14.30	Oxidative phosphorylation	21	C11ORF48, DDT, RPS21, NDUFA11	4.7%	MS, HIV, RSV, JDM	SOJIA, Pregnancy, Burkholderia
M14.31	Cell cycle	22	AP3M2, ANAPC4, NAT9, PFAAP5	5.0%		TB, SOJIA, Transplant,

						Staph, Burkholderia
M14.48	Inflammation	17	CTSS, DPEP2, NUP214, FCGRT	4.4%	Staph	HIV
M14.53	Erythrocytes	16	IGF2BP2, CHPT1, CDC34, RBM38	11.2%	Staph, RSV, Transplant, Melanoma	HIV
M14.67	Gene Transcription	16	EAF1, C18ORF32, FLI1, SAP30L	4.2%	HIV	
M14.76	Leukocyte activation	15	CD93, EIF2C4, FAM8A1, PECAM1	4.9%	Staph, Burkholderia, SOJIA	MS
M15.5	TBD	53	COPB1, GNG2, TAF7, SMEK2	2.7%		
M15.39	TBD	32	IKBKG, AP1M1, MGC3731, RAB40C	5.1%		
M15.42	TBD	30	HIVEP2, GOPC, SNW1, TMEM199	3.7%	HIV	Transplant
M15.49	TBD	27	RCOR3, CCNK, UBL3, PPP3CB	2.7%		Transplant
M15.55	Protein phosphorylation	24	DMXL2, HERC3, KIF5B, NLRC5	3.9%	SOJIA	
M15.61	Monocytes	25	KCNMB1, ANKRD57, SLC27A1, ZFHX3	7.4%		Burkholderia
M15.67	TBD	22	C19ORF56, C1ORF144, SPSB3, HMG20B	3.7%		
M15.86	Interferon	15	GALM, KIAA1618, MOV10, TIMM10	11.4%	HIV, Burkholderia, SOJIA, RSV, Influenza, SLE, TB	
M15.97	TBD	17	C12ORF10, SERF2, SH3GLB2, ROGDI	4.7%		
M15.102	Prostanoids	15	IL5RA, GPR44, OLIG2, PRSS33	18.4%		
M15.125	TBD	15	SIPA1L3, HUWE1, PSMD5, TEX261	4.2%		
M16.18	TBD	82	C11ORF31, DISP1, SALL2, ZNF543	3.9%		
M16.66	TBD	34	MC1R, KRI1, ZNF248, SCNN1D	5.5%		TB
M16.108	TBD	16	FMO5, ASRGL1, PI4K2A, SPIN3	4.3%		

Listed are the 32 modules used in the TFA assay and their summary annotations. The size of the original modules is noted and the 4 central genes selected for the assay and listed. The module variability in healthy controls was determined from longitudinal data collected every 2 weeks for

~30 weeks from 18 healthy controls. The disease activity notes the behavior of these modules among the 16 disease datasets that were used for module construction.

TFA primers

Listed are the primers used for the 8 housekeeping and 132 TFA genes.

Module	refSeq	Gene symbol	Forward Primer	Reverse Primer
House keeping	NM_004946.1	DOCK2	GTGTGCTCCACAAAGCTCAC	CCTGTAGCAGTTGAGGCTTCATA
House keeping	NM_004048.2	EEF1A1	ACTGGGCAGTGAAAGTTGAC	CCCTTCCACTCATAGGGTGTA
House keeping	NM_001004.3	FAM105B	CTGAGGAACCACCTCAACCA	CACAGCATAGGCAAGAAGGAAC
House keeping	NM_000146.3	FTL	AAACCCCAGACGCCATGAA	AGGAAGTGAGTCTCCAGGAA
House keeping	NM_033546.3	MYL12B	ACAGAGATGGCTTCATCGACAA	ACATGGTCAGGAACATGGTGAA
House keeping	NM_021019.4	MYL6	GACTTCACCGAAGACCAGAC	TGTACAGGATCTTGCCATCAC
House keeping	NM_001014.3	RPS10	GGTCAGCCACCTCAGTAAAA	TTGGCTGTAAGTTTATTCAATGCA
House keeping	NM_001028.2	RPS25	GGTGTCTGCTGCTATTCTCC	TCTTTCTTGGCCGACTTTCC
M10.2	NM_000518.4	HBB	AAGTGCTCGGTGCCTTTAGT	AGCTCACTCAGTGTGGCAAA
M10.2	NM_021706.2	LAIR1	CGTCGGACAACAGTCACAA	AGAGGAAGACCACTGAGACC
M10.2	NM_004152.2	OAZ1	CGAGCCGACCATGTCTTCA	AAGCTGAAGGTTCGGAGCAA
M10.2	NM_001016.3	RPS12	GGAGAATGGGTAGGCCTTTGTA	TGACATCCTTGGCCTGAGAC
M10.4	NM_002483.3	CEACAM6	AGATTGCATGTCCCCTGGAA	GGGTGGGTTCCAGAAGGTTA
M10.4	NM_001816.2	CEACAM8	CCCATCTCAGCCCCTTCC	CGGGTTCCAGAAGGTGAAAA
M10.4	NM_001925.1	DEFA4	TGGGATAAAAGCTCTGCTCTTCA	TGTTCGCCGGCAGAATACTA
M10.4	NM_001972.2	ELANE	TCTGCCGTCGCAGCAA	TTAGCCCGTTGCAGACCAA
M12.11	NM_005053.2	RAD23A	GATCCGCATGGAGCCTGAC	AGGCATCACGACCCTTCTCA
M12.11	NM_144563.2	RPIA	ACTTCGAATGGCTGTCAACA	CGGTCAAACTTCCAGTCCAA
M12.11	NM_005067.5	SIAH2	CAGGAACCTGGCTATGGAGAA	GTGCAGGGTCAGGGAACA
M12.11	NM_015049.1	TRAK2	CCTACTGCTCATCTCTACTTCTCC	CGCATAGTCCCCTCAATCTCA
M12.15	NM_001775.2	CD38	ACCTCACATGGTGTGGTGAA	GTTGCTGCAGTCCTTTCTCC
M12.15	NM_016459.3	MGC29506	AGCTGAGCGAGTTGGTCTAC	TTGGTCCACTTCTCGAACTCC
M12.15	NM_001192.2	TNFRSF17	TCTTTGGCAGTTTTCGTGCTAA	TTAGCCATGCCCAGGAGAC
M12.15	NM_001071.1	TYMS	TGTCGGTATTCGGCATGCA	CAGAACACACGTTTGGTTGTCA
M12.2	NM_000690.2	ALDH2	GGGAAGAGCCCCAACATCA	TGGCCCTGGTTGAAGAACA
M12.2	NM_004364.2	CEBPA	CATCGACATCAGCGCCTACA	CCGGCTGTGCTGGAACA
M12.2	NM_032048.2	EMILIN2	GTGAACGACGGGGATGTTTA	ATCAGGTAGCGCCCATCATA
M12.2	NM_003937.2	KYNU	GCCATCACAAAAGCTGGACAA	TCCAACTGCATGTGCTAGATCA
M12.3	NM_018091.4	ELP3	TGCCAAACGTGGGACTAGAA	AGCCCATCGGGACGAAAA
M12.3	NM_006055.1	LANCL1	GGATGCAGAACACCAGACAC	AGCAGGTCAGCCAGGAAATA
M12.3	NM_015231.1	NUP160	GTGGCAAGTTGTTCTCCGTAA	GACTCCTCCATCAGCTCCAA
M12.3	NM_004623.2	TTC4	GCACCTTGCTACAGGTTCTA	AGAGGATCCTACACAGACCAA
M12.4	NM_014047.2	C19ORF53	TGCAGCAGCAAAAGCTCAA	ACGTCATGTTCGATCTTCTTCC
M12.4	NM_144716.2	CCDC12	AGCCAAAGACCAAGCATCTCA	GGACATAGTTCCGCAGCCTAA
M12.4	NM_004424.3	E4F1	ACCAAGGGCTCACTCATCC	TCCACACTTGGAGCACTTGTA
M12.4	NM_014222.2	NDUFA8	AAGATCCGAGGCGGTGTTTA	GCTCTGCACAGTGACGTTTTA
M12.5	NM_052857.2	CCDC16	TGGACAAAGAGTGGGACGAA	CCTCATCCTCTTCGGCAACTA
M12.5	NM_018142.2	INTS10	AGCTGGGAGTTGCTCTATTCC	CAAAGCCAAGTATCCGTCTTCC
M12.5	NM_003161.2	RPS6KB1	TTTATTGGCAGCCCACGAAC	GAAGCACCTCTTCCCCAGAA

M12.5	NM_022340.2	ZFYVE20	CACACACCTGACATCGTGAA	TTCTGGAGCTTTCTGGTCAAC
M13.16	NM_025144.2	ALPK1	TGGTGCAGTCGGTCTGTATA	TTAACTCTGCTGCTTCGTACC
M13.16	NM_000167.3	GK	AAGCCAAGAATCTCTTTCAGTTCA	GCAGAGGTAATTCTGTCATCCC
M13.16	NM_000891.2	KCNJ2	AACGGTACCTCGCAGACATC	ACGAAAGCCAGGCAGAAGAT
M13.16	NM_052972.2	LRG1	ATGTCCTCTTGGAGCAGAC	GCAGCAGGAACAGAGTTCTA
M13.19	NM_014810.3	CEP350	CGTGTGGAATTTCGTGAACC	CCAGATGGCTGGAACTGAAA
M13.19	NM_005923.3	MAP3K5	ATCATTCGGAAGGCGGTACA	ACTCTCAGATGCAAGGCTGAA
M13.19	NM_002649.2	PIK3CG	TGGATCTATGCCTCCTGCCATA	TGTCGTGGCGTCTTTCACAA
M13.19	NM_139283.1	PPTC7	GGCAACAGATGGACTCTTTGAC	TGCTTCTGGCAGTCTGTTGTA
M14.30	NM_024099.3	C11ORF48	GGCCTTCTTCCAAGACCAAA	CAGCTGAGTCCAGCGTTAAA
M14.30	NM_001355.3	DDT	TTTCTCACCAAGGAGCTAGCC	CCAATCTGCCAGGACTCCAA
M14.30	NM_175614.2	NDUFA11	AGCCTACAGCACCACCAGTA	CCCGGAGGATTGAGTGTGAC
M14.30	NM_001024.3	RPS21	CATCATCGGTGCCAAGGAC	GCCATTAAACCTGCCTGTGAC
M14.31	NM_013367.2	ANAPC4	AGTAACTCGGATGGCCAGAA	CTTCCCATGCTTCACACATACA
M14.31	NM_006803.2	AP3M2	ACACACATTCGACCCAGTCA	ATGGTCCCCTTCAAACTTGGTA
M14.31	NM_015654.3	NAT9	CTGACCCTGGAGCAGGAGTA	TCGGCATCCAGCACAATGAA
M14.31	NM_014887.1	PFAAP5	AGCTTGGGAAATGAAGCCATA	TTCCACCAAGTTTCAGGTTCA
M14.48	NM_004079.3	CTSS	CACCACTGGCATCTCTGGAA	GAGACGTCGTACTGCTTCTTCA
M14.48	NM_022355.1	DPEP2	GGTGTCTTTGTCCATGGGAGTA	ACAGCCTTGATGTGGTCGAA
M14.48	NM_004107.3	FCGRT	TCATCGGTGTCTTGCTACTCA	CTCCACGAAGGGAGATCCAA
M14.48	NM_005085.2	NUP214	CTGCGGCTTCAACAGAAGTTA	TCCTCCAGTAGCCAAGATTCC
M14.53	NM_004359.1	CDC34	CACCAGCCTTTCGGTTCC	GGGTGGAGGATGGAGATACAC
M14.53	NM_020244.2	CHPT1	CATGGTGGTGTTGGCAAGAA	ATGTGGAGTCCAGGTGACAA
M14.53	NM_006548.4	IGF2BP2	AGATTCGCCGGAGCCTCTA	TGATGACCATCCTTTCGCTGAC
M14.53	NM_183425.1	RBM38	GCAAGTCCCGCGGCTA	GCCCAGATATGCCAGGTTCA
M14.67	NM_001035005.2	C18ORF32	GCCATATATATACCCTCTGGTTTCC	TCTGTTGGTCCTTTTGTTGGTA
M14.67	NM_033083.6	EAF1	GGCAAAGGAGATGAAGTCACAA	TGTTCCCCTTGAACACAGTCA
M14.67	NM_002017.2	FLI1	ACACGGAAGTGCTGTTGTCA	GTCGGTGTGGGAGGTTGTATTA
M14.67	NM_024632.4	SAP30L	GGCTTCAATAAGGCCCAGTTA	TCCGATTTCTGGTCCAGTCTA
M14.76	NM_012072.3	CD93	GACAGTTACTCCTGGGTTCCA	CTAGGGCCACCTCACTTTCA
M14.76	NM_017629.2	EIF2C4	AGTGGTTATCCTGCCTGGAA	ACACTGTGTGGCCATACCTA
M14.76	NM_016255.1	FAM8A1	GAAAATGATGGTTGTGGCACTTA	CCTGCTCCCCAAATGCA
M14.76	NM_000442.3	PECAM1	GCCCTCCACAAATGGGAAAAC	TGTGCGTTGCCTGAATGAAC
M15.102	NM_138983.1	GPR44	CCAGGGCTGGAATCCTGTG	GGCAGAGTGGCTTCAGTGT
M15.102	NM_000564.2	IL5RA	GATCAGCTGTTTGCCCTTCA	TTCCTTCAATCTCTGCTGTGAC
M15.102	NM_005806.2	OLIG2	CGGAGCGAGCTCCTCAAA	ATGGCCCCAGGGGAAGATA
M15.102	NM_152891.2	PRSS33	ACACCATGCCGGGTCAC	AGCAGCGGCACCCTTAC
M15.125	NM_031407.3	HUWE1	TTGCAGGTGTCTGCAAATCC	TGGAGTCCAACTGAAGGAGAC
M15.125	NM_005047.2	PSMD5	TTGTTGGGGCAGATTCAGAC	CATGACAGCCAGGTTTCCAA
M15.125	NM_002584.1	SIPA1L3	TACGTGAGATACAAGCCATCCC	GTGGCTGAAGTGGGGATCA
M15.125	NM_144582.2	TEX261	GAGAACGTCCTGCCCTCTAC	CCGCTTGCCTTTGGTGAAATA
M15.39	NM_032493.2	AP1M1	GGTCCGTTTCATGTGGATCA	GAAAGAAAAGACCAGCGACAC
M15.39	NM_001099856.1	IKBKG	GCTGCCTGGAGGAGAATCA	GGCTGGCTTGGAAATGCA
M15.39	NM_024313.1	MGC3731	CGCCACCAGGAATACTTGAA	ACTCCGTCTTTGCTGTCAC

M15.39	NM_021168.2	RAB40C	TCTGCACCATCTTCAGGTCCTA	CCAGCGGTTGGTGATGTCATA
M15.42	NM_020399.2	GOPC	AGCCCAGTCTGTGTCTCAAA	GCTTGGGTTTCTGTCAGTTCA
M15.42	NM_006734.3	HIVEP2	TCCAAACAAACTGAGCCAATCC	GGCCACGTCCTCTGACATATA
M15.42	NM_012245.2	SNW1	CCAAATTGGCAGAAGCCCTCTA	TACTTGGGCACGCATTTCCA
M15.42	NM_152464.1	TMEM199	CTGCCTTCGTCTGCACTTAC	GCGACGATCAATGCAGCTA
M15.49	NM_003858.3	CCNK	TTGGCCAGTTTGGAGATGAC	TCAAACTTGATGGTCTGCAGTAA
M15.49	NM_021132.1	PPP3CB	GTGTTCATGGTGGACTTTCAC	AATGCAGGTGGCTCTTTGAA
M15.49	NM_018254.2	RCOR3	GCAAAGGAAAAGCATGGCTA	GATCAGCAAGGGACTTCTCA
M15.49	NM_007106.2	UBL3	TCAAGGACGATTTCTACATGGAA	ATGTCTCTCTGGCCACCAA
M15.5	NM_016451.3	COPB1	CATTCCTGTTCTGTCCGATTTCC	CTGCTGCTTCGTTGTTGTCA
M15.5	NM_053064.3	GNG2	CAACAACACCGCCAGCATA	TGCTTCACAGTAGGCCATCA
M15.5	NM_020463.1	SMEK2	ACCTCAAAACAGGGATGCA	TCTGTAGCAGCTGATCTGAC
M15.5	NM_005642.2	TAF7	TCCAGGAATGTCTGGTCACA	CTGCTGCTGAGGTCATTGAA
M15.55	NM_015263.2	DMXL2	GAGGATAGAGAACGGGGTTTAC	ATGATCCTGTACCTGCCTCA
M15.55	NM_014606.1	HERC3	CCCTAAGAGAGCTGAGCATTCA	TCCACTGCTTCTTCACCATCA
M15.55	NM_004521.1	KIF5B	TCGCCAAGTTTCAGGGAGAA	GAGATGTGCTTGACTGGAACAC
M15.55	NM_032206.3	NLRC5	CCTGTGTCCACGGGTTAAAA	TCCTCCTCGTTGGATCTGAA
M15.61	NM_023016.3	ANKRD57	ACTTCAAGGAGCTGGTGAAC	TCTTCTTGAGGTGCACGTAC
M15.61	NM_004137.2	KCNMB1	GACTGTGCTGCCCCTCTAC	GGTCCCTGATGTTGGTCTCAA
M15.61	NM_198580.1	SLC27A1	CTGGGACGACTGCATCAAGTA	CGGCTGCTTCAGCAGGTA
M15.61	NM_006885.3	ZFHX3	GAGGCCAGCCTGAAGTTGTA	TGCACAGAACGCAGTGGTA
M15.67	NM_016145.2	C19ORF56	ACGCCGGACTACATGAACC	AGCTGATGAAGGAGCAGTAGAC
M15.67	NM_015609.2	C1ORF144	GATCTCCCAACCCGAAGACA	TTGAAGCCTTGAGACCCATCA
M15.67	NM_006339.1	HMG20B	TCGAGCTCTGGGCTCATGAA	TGGGAACATCGAAGGTGGAGAA
M15.67	NM_080861.3	SPSB3	GCTCCACTAACTGGGGCTAC	GCGGCAGCGTGGAGTA
M15.86	NM_138801.1	GALM	CTGTGGATGAAACCCTGATTCC	ACTGGCTTTCTCAGGTCGAA
M15.86	NM_020954.2	KIAA1618	AGGAAGCAGATGTCCAGGAA	AGAGAGATGATGGCGTGGAA
M15.86	NM_020963.2	MOV10	CAGCAAACACCACAAGTCAC	ACATCCACACCATGTTTCCC
M15.86	NM_012456.2	TIMM10	TGCGGAGCTGGAGGTG	GTAGTGAGGAGGCACACACT
M15.97	NM_021640.3	C12ORF10	TGCACGAGTTGCTCGACTTA	ATCCATTGCACGCTTGAACC
M15.97	NM_024589.1	ROGDI	AGGTGAAGGGTGTGCTGAC	TTGTTCCGGGGCATCTTCA
M15.97	NM_001018108.2	SERF2	CCGAGATGACGGGCTTTCT	GGGTTCCTCCTTCTTCTCGTT
M15.97	NM_020145.2	SH3GLB2	GCAAAGCGAGGCTGAAGAA	ACGAGGTCTAGTCTCCTGAAA
M16.108	NM_025080.2	ASRGL1	CGCAGTCCAGTGTATAGCAA	CCTTGGTCAGTCAGAAAGCA
M16.108	NM_001461.1	FMO5	TTTGCACTGGCCATCACAC	GTGGAAGTACTGCCCTTTGAAC
M16.108	NM_018425.2	PI4K2A	AACTCAACATTGTTCCCCGTAC	TCACTCGGTCAATGGCACTA
M16.108	NM_001010862.1	SPIN3	GAGAAGTCATAGACAGCCTGGT	ACCATGCCAGTTCTCTTGGA
M16.18	NM_170746.2	C11ORF31	CAGTAAAGGTGAACCCGACGAA	GCCCCTTCTTAATCCCAGTCC
M16.18	NM_032890.2	DISP1	TCTGTGCCTTGGTTGGAGTA	CTCTGGCCTATTGCTGTTCC
M16.18	NM_005407.1	SALL2	GGCAACTCGCATGCAACTAA	AGTGGACTTGAAGTGGTTGGT
M16.18	NM_213598.1	ZNF543	GACCTCTCCCAAAGCTCCTA	AGGCCAGGTGAGAAAAGGTA
M16.66	NM_023008.3	KRI1	AGATTCGGAACCCAGATTCCC	ATCCAACTCAGGGTCGTTCC
M16.66	NM_002386.2	MC1R	TGGTGAGCTTGGTGGAGAA	GCAGCAGATGAAGCAGTACA
M16.66	NM_002978.2	SCNN1D	CTGCTACACAACACCTCCTACA	CCACAGGAGCAGGTCTCC

M16.66	NM_021045.1	ZNF248	AAGGAGAAGAGCCCTGGATA	TTTCCTGGCTGCTCTCTAAC
M8.2	NM_005808.2	CTDSPL	CCGGTTGAAATCGATGGAACTA	CTGGCAGTAAAGAGCACACA
M8.2	NM_031469.2	SH3BGRL2	CCCCTGCCACCTCAGATATTTA	TGTGTTGCTTTCCTTGGATTCA
M8.2	NM_178562.2	TSPAN33	TCCTGTTGCTTGCCTACTCC	AGTAGTCAAAGGCCTGCATACC
M8.2	NM_006675.3	TSPAN9	GGGGCCATCAAGGAAAACAA	TCTGCTAGGAGGATGACCAAC
M8.3	NM_006417.3	IFI44	GGCTTTGGTGGGCACTAATA	TGCCATCTTTCCCGTCTCTA
M8.3	NM_005101.1	ISG15	CTGAGAGGCAGCGAACTCA	GCTCAGGGACACCTGGAA
M8.3	NM_002346.1	LY6E	TGCTCCGACCAGGACAACTA	GGCTGTGGCCAAATGTCAC
M8.3	NM_199139.1	XAF1	AGAGCAGAACATGGAAGGAGAC	AGGGTGAAGTTGGCAGAGAC

TFA Analysis Code

The R code below can be used to analyze TFA data that have been converted from Ct values to FC relative to reference samples. The structure expected for this analysis is a directory for the

```
# set up the project directory

currentDirectory<-"/Projects/Project_Name/"
dataDirectory<-paste(currentDirectory, "Data/", sep="")
resultsDirectory<-paste(currentDirectory, "Results/", sep="")
scriptsDirectory<-paste(currentDirectory, "scripts/", sep="")
source(paste(scriptsDirectory, "genomicCBCMATfunctions.R", sep=""))
# load required packages
packageLoad("limma")
# read in data
expData<-readData(dataDirectory, "FC.csv")
designData<-readData(dataDirectory, "designFile.csv")
moduleAssignData<-readData(dataDirectory, "moduleAssignFile.csv")
# FC values have row names in the first column
rownames(expData)<-as.character(expData[,1])
expData<-expData[,-1]
subExpData<-subsetExpData(expData, designData, moduleAssignData)
# next calculate FC values compared to controls (rather than reference samples)
FCcomparedToControls<-FCtocontrols(subExpData, designData)
# these are the individual module level results
FCmoduleLevelControls<-convertFCtoModule(FCcomparedToControls, moduleAssignData)
write.csv(FCmoduleLevelControls, file=paste(resultsDirectory, "individualModuleLevelFC.csv",
sep=""))
# next run gene level analysis using limma to find statistically significant genes
# use the original FC values (compared to reference sample) - get the exact same results if use
data compared to controls
geneLevel<-fitLimmaToGenes(subExpData, designData, moduleAssignData)
#compareGeneLevel<-fitLimmaToGenes(FCcomparedToControls, designData,
moduleAssignData)
write.csv(geneLevel, file=paste(resultsDirectory, "geneLevelStatistics.csv", sep=""),
row.names=FALSE)
# now calculate an average FC for each module
modLevelFCavg<-CaseVsControlModuleLevelFC(FCmoduleLevelControls, designData)
modLevelFCavg<-as.data.frame(modLevelFCavg)
colnames(modLevelFCavg)<-"FC"
write.csv(modLevelFCavg, file=paste(resultsDirectory, "groupModuleLevelFC.csv", sep=""))
```

The functions that are called above are defined below:

```
packageLoad = function(x) {
        # returns a matrix of the installed packages
        inst = installed.packages()
        matchVals = match(x, inst[,"Package"])
        missing = x[is.na(matchVals)]
        if (length(missing)>0)
        {
                cat("Installing packages...", "\n")
                lapply(missing,          install.packages,          repos="http://www.revolution-
computing.com/cran/")
        }
        if ("limma" %in% missing)
        {
                source("http://bioconductor.org/biocLite.R")
                biocLite("limma")
        }
        # then load libraries
        for (i in 1:length(x))
        {
                require(x[i], character.only=TRUE)
        }
}
readData<-function(path, dataFile)
{
        dataPath<-paste(path, dataFile, sep="")
        checkFile<-unlist(strsplit(dataFile, "\\."))
        fileType<-tolower(checkFile[length(checkFile)])
        if (fileType=="csv")
                curData<-read.csv(dataPath)
        if (fileType=="txt" | fileType=="tsv")
                curData<-read.delim(dataPath)
        return(curData)
}
# subset samples and genes to only those in design and moduleAssign
subsetExpData<-function(expData, designData, moduleAssignData)
{
        subExpData<-expData
        # first remove any columns (samples that are not in designData)
        matchIndex<-match(colnames(subExpData), designData$sampleID)
        if (any(is.na(matchIndex)))
        {
                remIndex<-which(is.na(matchIndex))
                subExpData<-subExpData[,-remIndex]
        }
```

```
    if (any(!(designData$groupID %in% c(0,1))))
    {
        # remove these samples that are not controls or cases (may be NTC or reference)
        keepIds<-designData$sampleID[c(which(designData$groupID==0),
which(designData$groupID==1))]
        # check - yes this is correct
        #table(designData$groupID[match(keepIds, designData$sampleID)])
        matchIndex<-match(keepIds, colnames(subExpData))
        keepIndex<-matchIndex[which(!is.na(matchIndex))]
        subExpData<-subExpData[,keepIndex]
        # check that this is correct - yes
        #table(designData$groupID[match(colnames(subExpData),
designData$sampleID)])
    }
    # next remove rows (such as housekeeping genes)
    matchIndex<-match(rownames(subExpData), moduleAssignData$geneID)
    if (any(is.na(matchIndex)))
    {
        remIndex<-which(is.na(matchIndex))
        subExpData<-subExpData[-remIndex,]
    }
    return(subExpData)
}

# calculate FC values compared to controls rather than reference samples
FCtocontrols<-function(subExpData, designData)
{
    controlIndex<-which(!is.na(match(colnames(subExpData),
designData$sampleID[which(designData$groupID==0)])))
    # convert to delta delta Ct values before standardizing to control means
    controlDelta<- -log2(subExpData[,controlIndex])
    controlMeans<-apply(controlDelta, 1, mean, na.rm=T)
    allDelta<- -log2(subExpData)
    allDeltaToControls<-apply(allDelta, 2, function(x)
    {
        x-controlMeans
    })
    allFCtoControls<-2^(-allDeltaToControls)
    return(allFCtoControls)
}

# convert the FC values from gene level to module level
convertFCtoModule<-function(FCvals, moduleAssignData)
{
    # loop through all modules
    uniModules<-sort(as.character(unique(moduleAssignData$module)))
    FCtoModule<-c()
    for (i in 1:length(uniModules))
    {
        curMod<-uniModules[i]
        curGeneIDs<-
moduleAssignData$geneID[which(curMod==moduleAssignData$module)]
        matchIndex<-match(curGeneIDs, rownames(FCvals))
        if (any(is.na(matchIndex)))
            matchIndex<-matchIndex[-which(is.na(matchIndex))]
        curMeans<-2^apply(log2(FCvals[matchIndex,]), 2, mean, na.rm=T)
        FCtoModule<-rbind(FCtoModule, curMeans)
    }
    rownames(FCtoModule)<-uniModules
    #apply(FCtoModule, 1, function(x) {sum(is.na(x))})
    return(FCtoModule)
}

# run model fits at the gene level
fitLimmaToGenes<-function(FCvals, designData, moduleAssignData)
{
    # make a design matrix - first order the columns in FCvals to match those in designData
    matchIndex<-match(colnames(FCvals), designData$sampleID)
    curGroups<-designData$groupID[matchIndex]
    if (all(curGroups %in% c(0,1)))
    {
        curMM<-model.matrix(~1+factor(curGroups))
        curFit<-lmFit(log2(FCvals), curMM)
        curFit<-eBayes(curFit)
        curTT<-topTable(curFit, coef=2, adjust.method="BH", number=nrow(FCvals))
        # output: gene name, module, test statistic, p-value, and adjusted p-value
        # need to add module to curTT
        curTT<-curTT[,1:5]
        curTT$module<-moduleAssignData$module[match(curTT$ID,
moduleAssignData$geneID)]
        curTT$geneName<-moduleAssignData$geneName[match(curTT$ID,
moduleAssignData$geneID)]
        return(curTT)
    }
    else
        print("ERROR: Columns in FC matrix belong to a group that is not case or control.")
}
```

```
# assumes the FC values are already at the module level
# starts with module level FC and compares case to control
CaseVsControlModuleLevelFC<-function(FCvals, designData)
{
    controlMatch<-match(designData$sampleID[which(designData$groupID==0)],
colnames(FCvals))
    if (any(is.na(controlMatch)))
        controlMatch<-controlMatch[-which(is.na(controlMatch))]
    controlVals<-FCvals[,controlMatch]
    caseMatch<-match(designData$sampleID[which(designData$groupID==1)],
colnames(FCvals))
    if (any(is.na(caseMatch)))
        caseMatch<-caseMatch[-which(is.na(caseMatch))]
    caseVals<-FCvals[,caseMatch]
    # take mean on log2 scale
    controlMeans<-2^apply(log2(controlVals), 1, mean, na.rm=T)
    caseMeans<-2^apply(log2(caseVals), 1, mean, na.rm=T)
    modMeans<-caseMeans/controlMeans
    return(modMeans)
}
```

REFERENCES

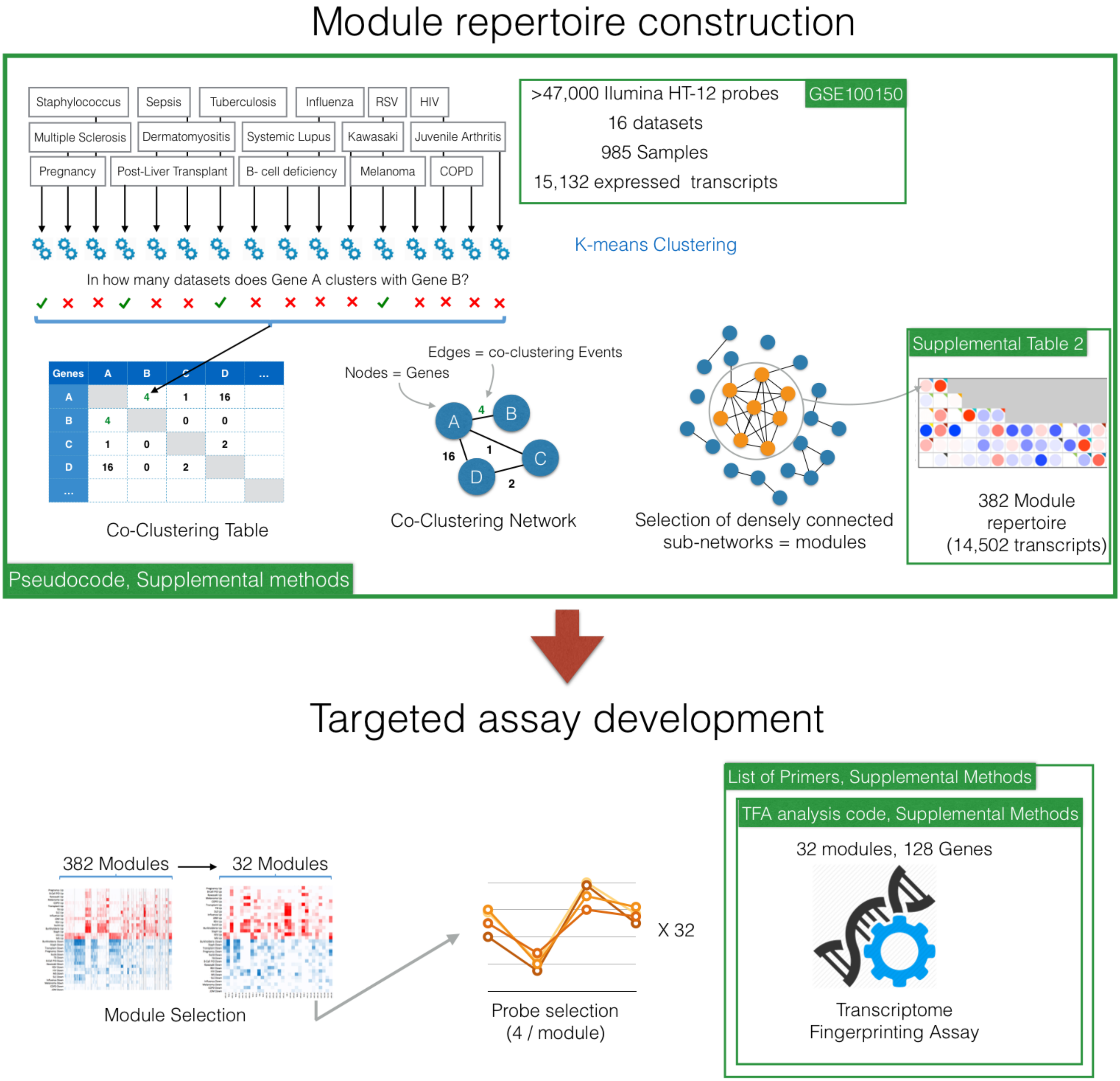
1. Hansson, G.K. and A. Hermansson, *The immune system in atherosclerosis.* Nat Immunol, 2011. **12**(3): p. 204-212.
2. Kanneganti, T.-D. and V.D. Dixit, *Immunological complications of obesity.* Nat Immunol, 2012. **13**(8): p. 707-712.
3. Heneka, M.T., M.P. Kummer, and E. Latz, *Innate immune activation in neurodegenerative disease.* Nat Rev Immunol, 2014. **14**(7): p. 463-477.
4. Chen, D.S. and I. Mellman, *Oncology meets immunology: the cancer-immunity cycle.* Immunity, 2013. **39**(1): p. 1-10.
5. Shearer, W.T. and C.G. Fathman, *30. Defining the spectrum of clinical immunology.* J Allergy Clin Immunol, 2003. **111**(2 Suppl): p. S766-73.
6. Singh, H., A.A. Khan, and A.R. Dinner, *Gene regulatory networks in the immune system.* Trends in Immunology, 2014. **35**(5): p. 211-218.
7. Shurin, M.R. and Y.S. Smolkin, *Immune-mediated diseases: where do we stand?* Adv Exp Med Biol, 2007. **601**: p. 3-12.
8. Vodovotz, Y., et al., *Solving Immunology?* Trends in Immunology, 2016.
9. Hoffmann, G.W., *A theory of regulation and self-nonself discrimination in an immune network.* European Journal of Immunology, 1975. **5**(9): p. 638-647.
10. Kidd, B.A., et al., *Unifying immunology with informatics and multiscale biology.* Nat Immunol, 2014. **15**(2): p. 118-27.
11. Altman, M.C., *Construction of a Third-Generation Transcriptomic Modular Repertoire in Clinical Immunology.* Nat Immunol, 2018.
12. Sugar, C.A. and G.M. James, *Finding the Number of Clusters in a Dataset.* Journal of the American Statistical Association, 2003. **98**(463): p. 750-763.
13. Chiche, L., et al., *Modular transcriptional repertoire analyses of adults with systemic lupus erythematosus reveal distinct type I and type II interferon signatures.* Arthritis Rheumatol, 2014. **66**(6): p. 1583-95.
14. Bombardier, C., et al., *Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE.* Arthritis Rheum, 1992. **35**(6): p. 630-40.
15. Kavanaugh, A.F., D.H. Solomon, and G. American College of Rheumatology Ad Hoc Committee on Immunologic Testing, *Guidelines for immunologic laboratory testing in the rheumatic diseases: anti-DNA antibody tests.* Arthritis Rheum, 2002. **47**(5): p. 546-55.
16. Summers, C.M., et al., *Functional polymorphisms of folate-metabolizing enzymes in relation to homocysteine concentrations in systemic lupus erythematosus.* J Rheumatol, 2008. **35**(11): p. 2179-86.
17. Becker, A.M., et al., *SLE peripheral blood B cell, T cell and myeloid cell transcriptomes display unique profiles and each subset contributes to the interferon signature.* PLoS One, 2013. **8**(6): p. e67003.
18. Gross, J.A., et al., *TACI and BCMA are receptors for a TNF homologue implicated in B-cell autoimmune disease.* Nature, 2000. **404**(6781): p. 995-9.

19. Rosenbaum, M., et al., *MZB1 is a GRP94 cochaperone that enables proper immunoglobulin heavy chain biosynthesis upon ER stress.* Genes Dev, 2014. **28**(11): p. 1165-78.
20. Crow, M.K., *Type I Interferon in the Pathogenesis of Lupus.* The Journal of Immunology, 2014. **192**(12): p. 5459-5468.
21. Obermoser, G. and V. Pascual, *The interferon-alpha signature of systemic lupus erythematosus.* Lupus, 2010. **19**(9): p. 1012-9.
22. Zickert, A., et al., *Interferon (IFN)-lambda is a potential mediator in lupus nephritis.* Lupus Sci Med, 2016. **3**(1): p. e000170.
23. Skubitz, K.M. and A.P. Skubitz, *Interdependency of CEACAM-1, -3, -6, and -8 induced human neutrophil adhesion to endothelial cells.* J Transl Med, 2008. **6**: p. 78.
24. Kaplan, M.J., *Neutrophils in the pathogenesis and manifestations of SLE.* Nat Rev Rheumatol, 2011. **7**(12): p. 691-9.
25. Tsokos, G.C., et al., *New insights into the immunopathogenesis of systemic lupus erythematosus.* Nature Reviews Rheumatology, 2016. **12**(12): p. 716-730.
26. Smith, C.K. and M.J. Kaplan, *The role of neutrophils in the pathogenesis of systemic lupus erythematosus.* Current Opinion in Rheumatology, 2015. **27**(5): p. 448-453.
27. Novy, M.J. and G.C. Liggins, *Role of prostaglandins, prostacyclin, and thromboxanes in the physiologic control of the uterus and in parturition.* Semin Perinatol, 1980. **4**(1): p. 45-66.
28. Blazkova, J., et al., *Multicenter Systems Analysis of Human Blood Reveals Immature Neutrophils in Males and During Pregnancy.* J Immunol, 2017. **198**(6): p. 2479-2488.
29. Luppi, P., et al., *Monocytes are progressively activated in the circulation of pregnant women.* J Leukoc Biol, 2002. **72**(5): p. 874-84.
30. Novershtern, N., et al., *Densely interconnected transcriptional circuits control cell states in human hematopoiesis.* Cell, 2011. **144**(2): p. 296-309.
31. Chaussabel, D., et al., *A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus.* Immunity, 2008. **29**(1): p. 150-64.
32. Cnossen, J.S., et al., *Are tests for predicting pre-eclampsia good enough to make screening viable? A review of reviews and critical appraisal.* Acta Obstet Gynecol Scand, 2009. **88**(7): p. 758-65.
33. Conde-Agudelo, A., et al., *Novel biomarkers for the prediction of the spontaneous preterm birth phenotype: a systematic review and meta-analysis.* BJOG, 2011. **118**(9): p. 1042-54.
34. Gibb, W. and J.R. Challis, *Mechanisms of term and preterm birth.* J Obstet Gynaecol Can, 2002. **24**(11): p. 874-83.
35. Challis, J.R., et al., *Prostaglandins and mechanisms of preterm birth.* Reproduction, 2002. **124**(1): p. 1-17.
36. Pertynska-Marczewska, M., et al., *Profile of peripheral blood neutrophil cytokines in diabetes type 1 pregnant women and its correlation with selected parameters in the newborns.* Am J Reprod Immunol, 2010. **63**(2): p. 150-60.
37. Walsh, S.W., *Eicosanoids in preeclampsia.* Prostaglandins Leukot Essent Fatty Acids, 2004. **70**(2): p. 223-32.
38. Laresgoiti-Servitje, E., *A leading role for the immune system in the pathophysiology of preeclampsia.* J Leukoc Biol, 2013. **94**(2): p. 247-57.

39. Faas, M.M., F. Spaans, and P. De Vos, *Monocytes and macrophages in pregnancy and pre-eclampsia.* Front Immunol, 2014. **5**: p. 298.

40. Speake, C., et al., *Longitudinal Monitoring of Gene Expression in Ultra-Low Volume Blood Samples Self-Collected at Home.* Clin Exp Immunol, 2016.

41. Bates, D., et al., *Fitting Linear Mixed-Effects Models Using lme4.* Journal of Statistical Software, 2015. **67**(1).


Module repertoire construction
Staphylococcus
Sepsis
Tuberculosis
Influenza
RSV
HIV
Multiple Sclerosis
Dermatomyositis
Systemic Lupus
Kawasaki
Juvenile Arthritis
Pregnancy
Post-Liver Transplant
B- cell deficiency
Melanoma
COPD
>47,000 Ilumina HT-12 probes
GSE100150
16 datasets
985 Samples
15,132 expressed transcripts
K-means Clustering
In how many datasets does Gene A clusters with Gene B?
✓ ✗ ✗ ✓ ✗ ✗ ✓ ✗ ✗ ✗ ✗ ✓ ✗ ✗ ✗ ✗
Edges = co-clustering Events
Nodes = Genes
A
B
C
D
4
1
16
2
Co-Clustering Network
Selection of densely connected sub-networks = modules
Supplemental Table 2
382 Module repertoire (14,502 transcripts)
Co-Clustering Table
Pseudocode, Supplemental methods
Targeted assay development
382 Modules
32 Modules
Module Selection
X 32
Probe selection (4 / module)
List of Primers, Supplemental Methods
TFA analysis code, Supplemental Methods
32 modules, 128 Genes
Transcriptome Fingerprinting Assay

Genes	A	B	C	D	...
A		4	1	16	
B	4		0	0	
C	1	0		2	
D	16	0	2		
...					

Figure 1

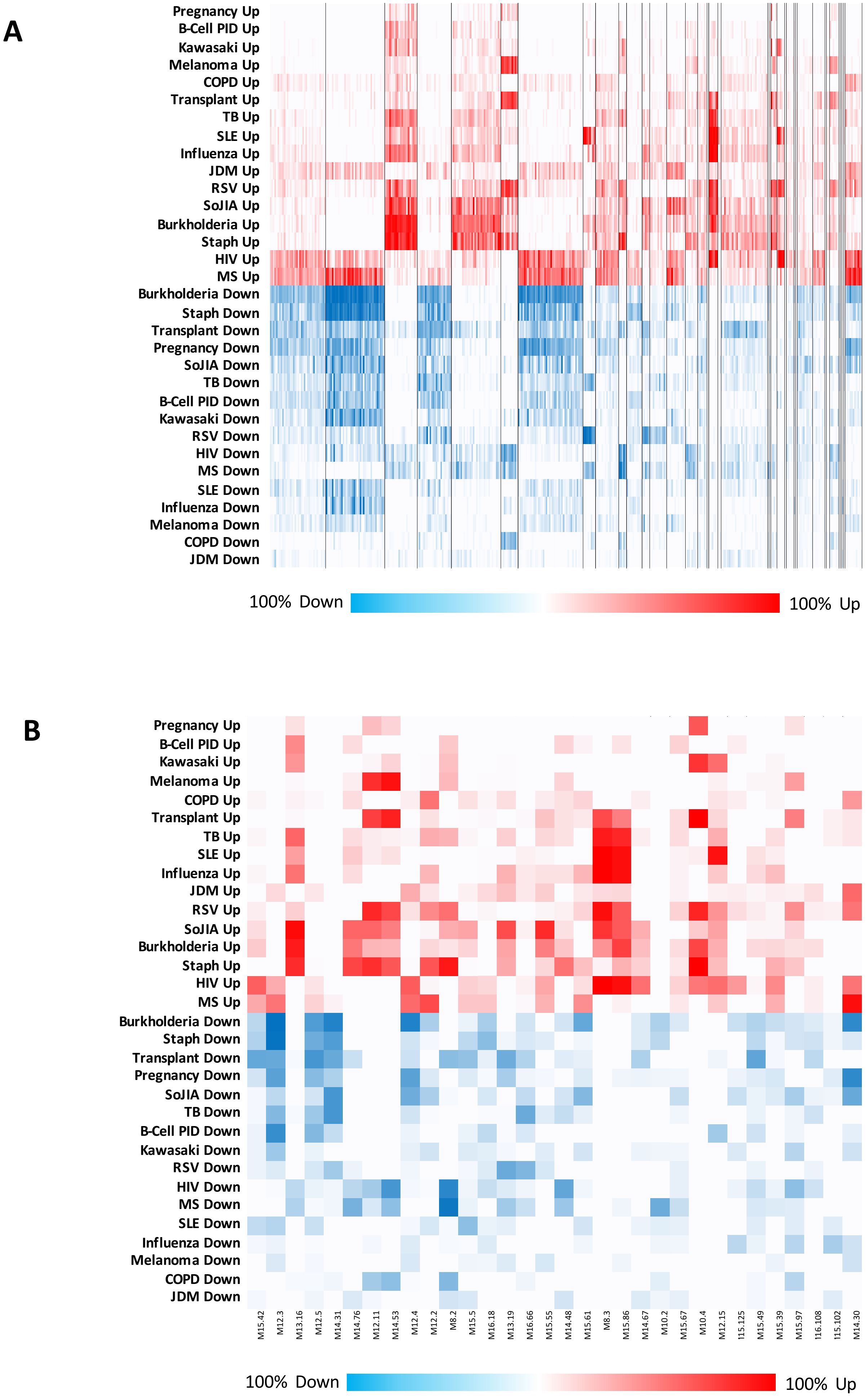

A
Pregnancy Up
B-Cell PID Up
Kawasaki Up
Melanoma Up
COPD Up
Transplant Up
TB Up
SLE Up
Influenza Up
JDM Up
RSV Up
SoJIA Up
Burkholderia Up
Staph Up
HIV Up
MS Up
Burkholderia Down
Staph Down
Transplant Down
Pregnancy Down
SoJIA Down
TB Down
B-Cell PID Down
Kawasaki Down
RSV Down
HIV Down
MS Down
SLE Down
Influenza Down
Melanoma Down
COPD Down
JDM Down
100% Down
100% Up
B
Pregnancy Up
B-Cell PID Up
Kawasaki Up
Melanoma Up
COPD Up
Transplant Up
TB Up
SLE Up
Influenza Up
JDM Up
RSV Up
SoJIA Up
Burkholderia Up
Staph Up
HIV Up
MS Up
Burkholderia Down
Staph Down
Transplant Down
Pregnancy Down
SoJIA Down
TB Down
B-Cell PID Down
Kawasaki Down
RSV Down
HIV Down
MS Down
SLE Down
Influenza Down
Melanoma Down
COPD Down
JDM Down
M15.42
M12.3
M13.16
M12.5
M14.31
M14.76
M12.11
M14.53
M12.4
M12.2
M8.2
M15.5
M16.18
M13.19
M16.66
M15.55
M14.48
M15.61
M8.3
M15.86
M14.67
M10.2
M15.67
M10.4
M12.15
15.125
M15.49
M15.39
M15.97
16.108
15.102
M14.30
100% Down
100% Up

Figure 2

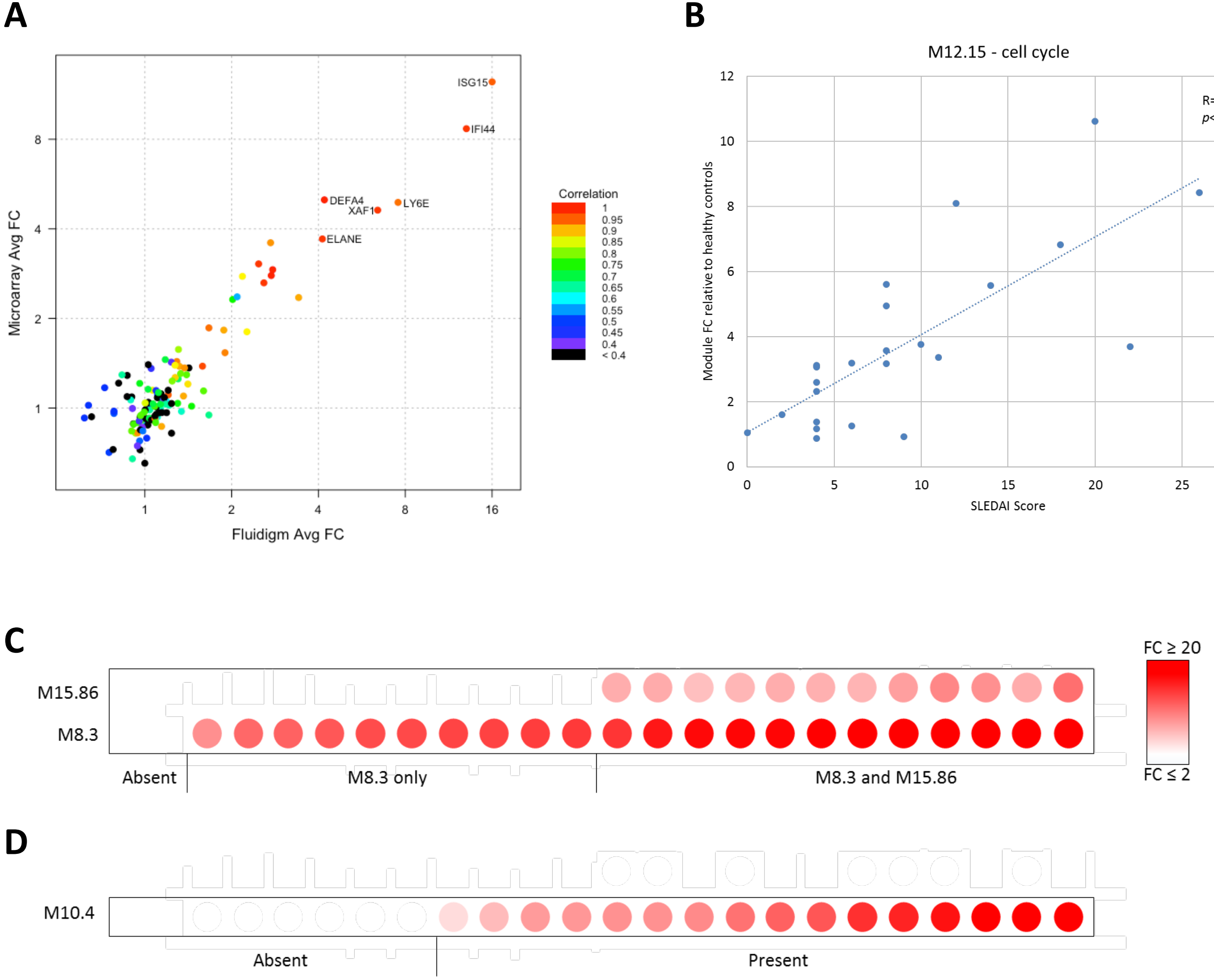
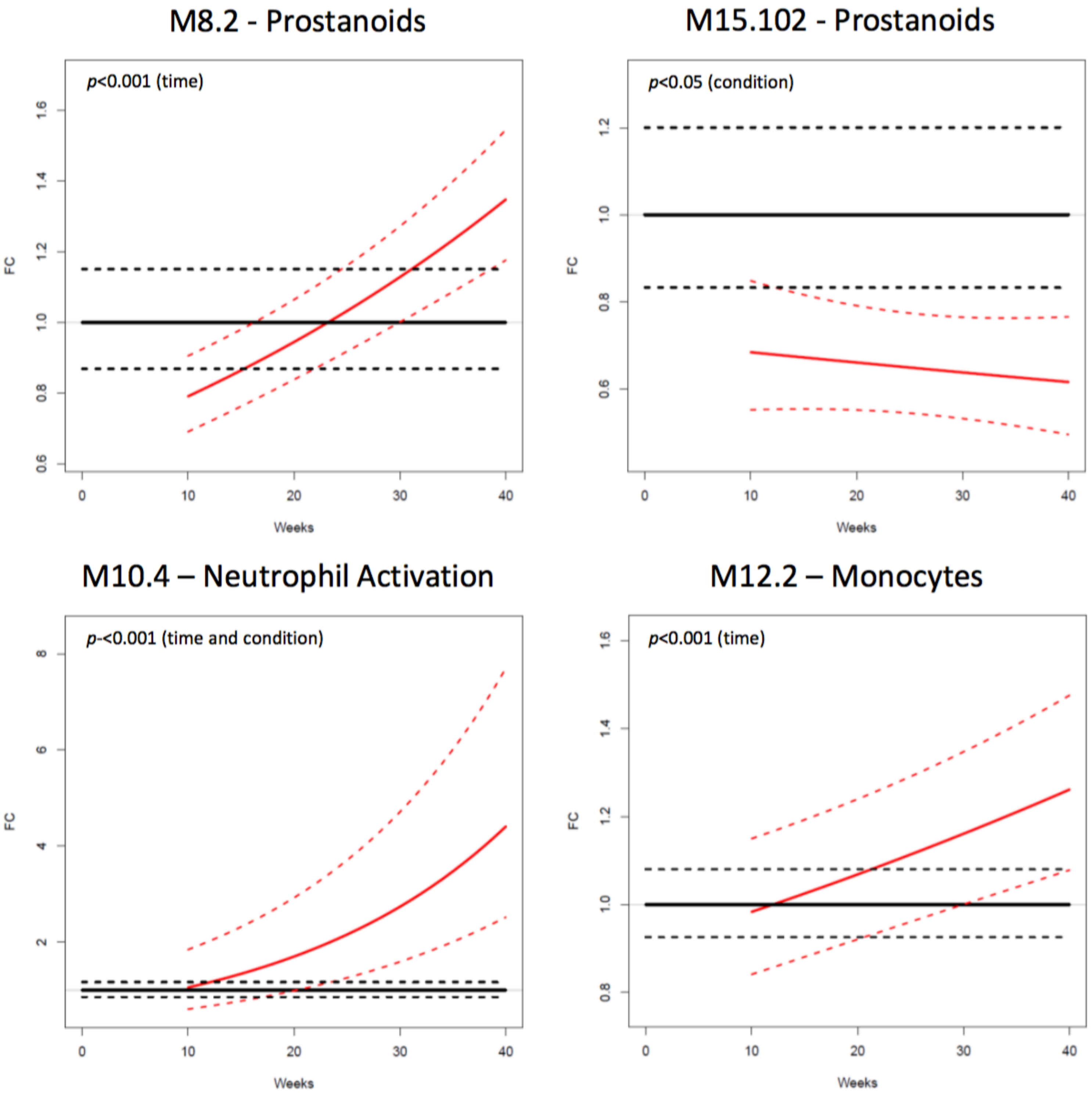

A
Microarray Avg FC
Fluidigm Avg FC
1
2
4
8
16
ISG15
IFI44
DEFA4
XAF1
LY6E
ELANE
Correlation
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
< 0.4
B
M12.15 - cell cycle
R=0.78
p<0.001
Module FC relative to healthy controls
SLEDAI Score
0
2
4
6
8
10
12
0
5
10
15
20
25
30
C
M15.86
M8.3
Absent
M8.3 only
M8.3 and M15.86
FC ≥ 20
FC ≤ 2
D
M10.4
Absent
Present

Figure 3



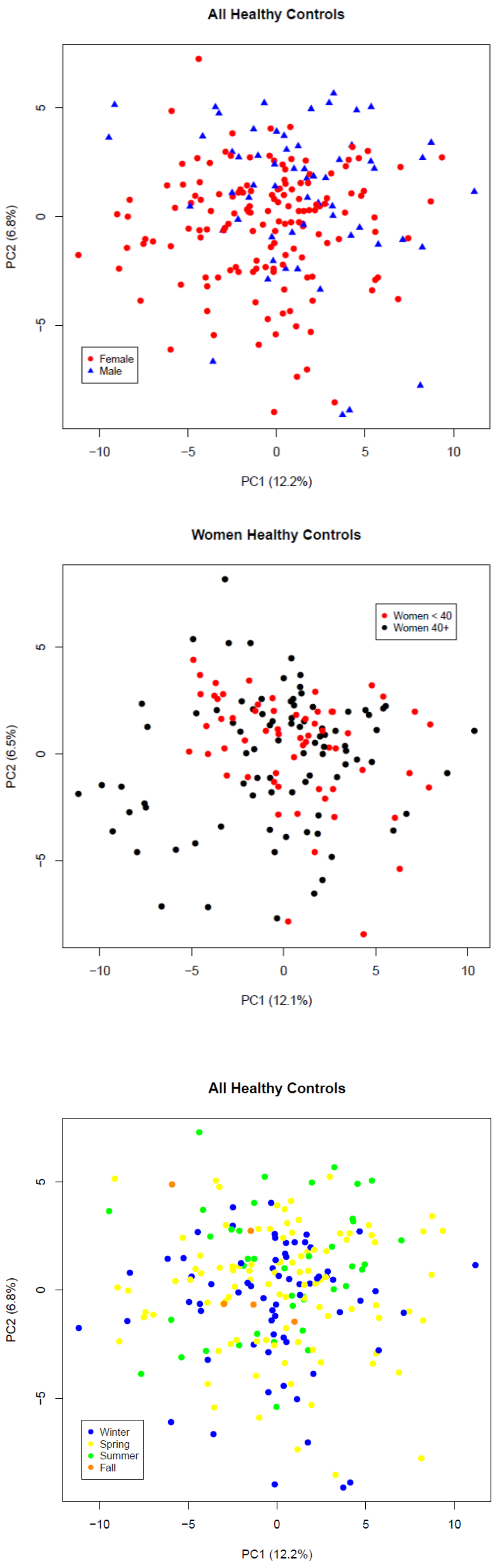
M8.2 - Prostanoids

M15.102 - Prostanoids

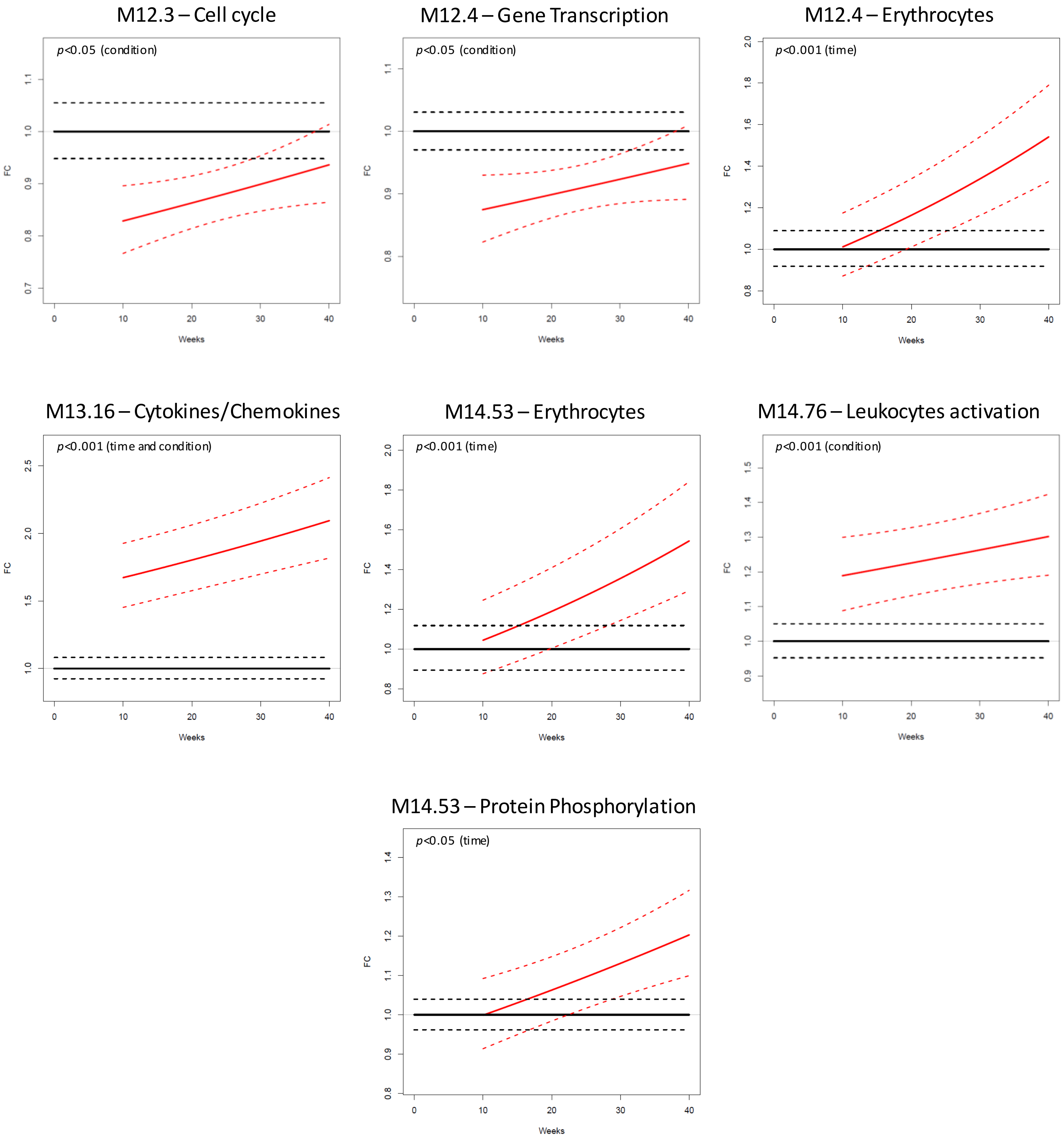
M10.4 – Neutrophil Activation

M12.2 – Monocytes

Figure 4



Supplementary Figure 1



M12.3 – Cell cycle

M12.4 – Gene Transcription

M12.4 – Erythrocytes

M13.16 – Cytokines/Chemokines

M14.53 – Erythrocytes

M14.76 – Leukocytes activation

M14.53 – Protein Phosphorylation

Supplementary Figure 2