Peripheral blood gene expression profiles in metabolic syndrome, coronary artery disease and type 2 diabetes

Britney L. Grayson[1], **Lily Wang**[2], and **Thomas M. Aune**[1,3]

[1]Department of Microbiology and Immunology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

[2]Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

[3]Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

Abstract

To determine if individuals with metabolic disorders possess unique gene expression profiles, we compared transcript levels in peripheral blood from patients with coronary artery disease, type 2 diabetes and their precursor state, metabolic syndrome to those of control subjects and subjects with rheumatoid arthritis. The gene expression profile of each metabolic state was distinguishable from controls and correlated with other metabolic states more than with rheumatoid arthritis. Of note, subjects in the metabolic cohorts over-expressed gene sets that participate in the innate immune response. Genes involved in activation of the pro-inflammatory transcription factor, NF-κB, were over-expressed in coronary artery disease while genes differentially expressed in type 2 diabetes play key roles in T cell activation and signaling. RT-PCR validation confirmed microarray results. Furthermore, several genes differentially expressed in human metabolic disorders have been previously shown to participate in inflammatory responses in murine models of obesity and Type 2 diabetes. Taken together, these data demonstrate that peripheral blood from individuals with metabolic disorders display overlapping and non-overlapping patterns of gene expression indicative of unique, underlying immune processes.

Keywords

Coronary artery disease; Gene expression profiles; Gene set; Metabolic syndrome; Rheumatoid arthritis; Type 2 diabetes

Corresponding Author: Thomas.M.Aune@Vanderbilt.edu, 1161 21st Avenue South, MCN T-3219, Nashville, TN 37232, USA, office: 615-343-7353, fax: 615-322-6248 .

Introduction

Type 2 diabetes (T2D) is a metabolic disorder of peripheral insulin resistance resulting in hyperglycemia and ultimately decreased insulin secretion from the pancreas. Risk factors for T2D include obesity, physical inactivity and family history of metabolic disorders or atherosclerosis[1]. Diabetes currently affects 6.3% of the United States population and approximately 90% of these cases are non-insulin dependent, or T2D[2]. Coronary artery disease (CAD) results from atherosclerotic plaque development in coronary arteries. These fibrous, fatty deposits can ultimately block the flow of blood resulting in angina and/or myocardial infarction. Hyperlipidemia predisposes to the development of these plaques, making obesity and physical inactivity also risk factors for CAD[3]. The prevalence of CAD in the United States is 4.1% [4]. Metabolic syndrome (MetS #0) is a precursor state to both T2D and CAD[5-13]. The International Diabetes Federation (IDF) defines this pre-disease state as central obesity plus any 2 of the following 4 characteristics: hypertriglyceridemia, low high-density lipoprotein (HDL) cholesterol, hypertension or raised fasting plasma glucose. The prevalence of MetS in the United States is as high as 39% using the IDF criteria. Diagnosis of MetS confers a 1.5-2.6 relative risk of developing CAD and a 3.5-7.5 relative risk of developing T2D. Additionally, the Framingham study determined that a portion of these relative risks persist even in the absence of obesity. This trio of disorders poses a significant threat to public health in the United States.

Inflammatory processes are involved in the pathogeneses of T2D and CAD. Visceral adipose tissue, present in abundance in many patients with T2D, produces inflammatory cytokines like IL-6 and TNF-α that are known to aid in the impairment of insulin signaling in adipocytes. These cytokines can activate a systemic immune response and recruit inflammatory cells, like lymphocytes, to visceral adipose tissue[14]. In the case of CAD, the lesion is not visceral adipose tissue, but rather fatty deposits in the vasculature. These deposits contain fat-laden macrophages and immunoreactive T-cells[3].

Gene expression profiling of blood or tissue samples is one way to assess cellular changes due to cell differentiation and aging[15; 16], disease pathogenesis[17-19] or pharmacological responses[20; 21]. One example of this is tumor typing; gene expression signatures are presently used to classify tumor types in breast cancer biopsies. This method can also be used to assess changes in peripheral whole blood of patients with common, complex diseases[22-24]. Individuals with autoimmune diseases [type 1 diabetes, multiple sclerosis, systemic lupus erythematosus and rheumatoid arthritis (RA)] display unique gene expression signatures in peripheral whole blood. Portions of these signatures are expressed in first degree unaffected relatives[25] however, disease-specific signatures are also found in peripheral blood and are sufficient to distinguish individuals with disease from control individuals[26]. Moreover, peripheral blood gene expression profiling can give insight into disease processes and suggest specific functional defects in cells. For example, peripheral blood gene expression in patients with RA contains low transcript levels of the tumor suppressor protein, p53. Consequently, T cells from patients with RA are resistant to gamma-radiation induced apoptosis, a p53 dependent pathway[27]. Gene expression profiling may also aid in diagnosing patients who have these often difficult to diagnose diseases; therefore, analysis of peripheral blood gene expression represents one approach to assessing

immune system changes, predicting cellular defects and diagnosing patients with immune-related disease in a minimally invasive way.

T2D, CAD and their precursor, MetS are not autoimmune diseases but feature inflammation as a possible pathogenic component. The purpose of our studies was to assess if these diseases also possess unique peripheral blood gene expression profiles and if so, what do the profiles indicate about the relationships among MetS, CAD and T2D. To address this question, we compared profiles of each disease state to control (CTRL) subjects, to an autoimmune disease, RA, and to each other.

Results

Peripheral blood gene expression profiling using microarrays has been shown sufficient to distinguish between phenotypically distinct cohorts of patients[22-24]. We sought to determine if subjects with MetS, CAD or T2D also possessed a gene expression signature in blood sufficient to distinguish these subjects from CTRL subjects and, if so, did this signature bear any resemblance to the signature of an autoimmune disease, RA. To do so, we recruited subjects with MetS, CAD and T2D (n=6, n=6, n=8, respectively), 6 subjects with RA, and 9 subjects who had never been diagnosed with a chronic illness, and were not presently taking medications for any diagnosed state, to serve as the CTRL cohort.

We analyzed all 35 peripheral blood samples for gene expression using the human exonic evidence-based oligonucleotide (HEEBO) array. Next, we normalized the data to a sum total intensity of 10,000, giving an average intensity per oligonucleotide probe of 0.2. Genes, with an average intensity of greater than 0.2, were used as data points for clustering analysis. The intensity values of the filtered set of genes for each array were inputted into The Institute for Genomic Research's multi-experiment viewer.

Initially the samples were imported in sets of control plus one disease or pre-disease state (RA, MetS, CAD or T2D, respectively) and then all samples were imported together. Using the Support Tree function, we created a dendrogram based on unsupervised clustering of each group of samples by similarity (Fig. 1). In other words, all samples were considered as 1 group for analysis even though they represented several different cohorts. Comparing CTRL and RA in an unsupervised manner created a dendrogram that gave 100% support to a separation of 2 groups, however, 3 CTRL samples were branched with the entire RA cohort. The unsupervised dendrograms of CTRL + MetS, CTRL + CAD and CTRL + T2D show similar trends towards clustering of the cohorts away from each other, but without 100% support. When all samples were imported in one analysis, we see loose clustering of the RA, T2D, CTRL and CAD cohorts, represented by those samples being placed in close proximity on the dendrogram. By clustering samples in an unsupervised analysis, the data suggest differences in gene expression between disease cohorts and control and also patterns that can be seen when all samples are considered together.

To further investigate differences in the gene expression profiles, we compared the RA, MetS, CAD and T2D groups individually to the CTRL cohort using a supervised significance analysis for microarray function. Each list of significantly differentially

expressed genes was used to run a bootstrap hierarchical clustering to determine the similarity of the patient samples to each other within each disease and their similarity to CTRL, which was known to be a separate group in the algorithm (Fig. 2). Although these were small sample sizes, the main branches of each support tree (denoted in black) represent 100% statistical support based upon bootstrap hierarchical clustering analyses. A black bar separates two main branches neatly clustering the RA cohort away from the CTRL cohort based on gene expression. For the T2D cohort, two T2D patients clustered with the CTRL group and conversely, the same 2 CTRL patients clustered on a branch with the MetS and CAD groups in their individual comparisons with CTRL. These separations indicate that the majority of subjects with RA, MetS, CAD or T2D are more like each other than the CTRL. Hierarchical clusters confirmed that expression of genes in peripheral whole blood was sufficient to distinguish between the autoimmune disease, RA, and CTRL, as well as the inflammatory metabolic states of MetS, CAD and T2D and CTRL. Further similarities and differences can be seen amongst the disease-affected subjects. In the RA group at least one further branch with 100% support was seen, indicating that gene expression is not entirely homogenous within this group.

Additionally, when all samples were analyzed together, the support tree indicates that 8/9 CTRL subjects cluster, with 100% support, on one branch with 2 T2D subjects, one of those being patient T2D 03 who previously clustered on the CTRL branch in the CTRL:T2D analysis (Fig. 2). The second branch features just one CTRL patient and the remainder of the patients in the disease cohorts. Five of the 6 RA patients cluster together on this branch, indicating that the RA signature is more like that of the metabolic diseases than CTRL; however, the RA patients are more like each other than the metabolic cohort patients. Furthermore, the remaining metabolic disease patients did not cluster in any particular pattern suggesting similarity amongst the MetS, CAD and T2D peripheral blood gene expression profiles. Taken together, this analysis demonstrates that subjects with MetS, CAD or T2D each possess a common gene expression signature in blood sufficient to distinguish them from CTRL and that these signatures may have overlapping components.

One possible source of differential gene expression in leukocytes is an alteration in the underlying genetic code[28]. Extensive genome wide analyses have been performed in RA, CAD and T2D revealing a number of single nucleotide polymorphisms (SNPs) associated with each individual disease. We probed our expression dataset to determine if genes associated with these SNPs showed differential expression in peripheral blood of subjects with disease versus CTRL subjects in any of our cohorts. A list of SNPs associated with RA, CAD or T2D was populated from The National Human Genome Research Institute[29] and a recent pathway based SNP analysis by Torkamani, *et al.*[30]. For the SNPs present in gene coding regions (45 SNPs associated with RA, 61 SNPs associated with T2D and 25 SNPs associated with CAD), we calculated expression levels of the encoded gene as an average for the RA, CAD and T2D groups. Each set of genes was analyzed for expression in disease groups and we found a number of correlations between a SNP, its encoded gene and differential expression of that gene. Eight genes with a disease-associated SNP were differentially expressed in the corresponding disease group-*CD244, IL2RA, PRKCA, SLC22A4* and *TRAF1* in RA, and *ADAMTS9, ANXA11* and *KCNQ1* in T2D (Table 1). *IL2RA*

and *TRAF1*, genes identified by SNP studies in RA, were also differentially expressed in T2D and the T2D SNP-identified genes *ADAMTS9* and *KCNQ1* were differentially expressed in RA. While SNPs are known to influence gene expression, we only found associations in RA and T2D, not CAD. Altered gene expression was not confined to just one disease state; differential expression of certain genes was shared between RA and T2D.

A second possible source of differential gene expression among individuals or disease groups was presence or absence of pharmacologic therapies. We performed this analysis and did not find differences in gene expression among the different subject groups associated with common therapies.

Because hierarchical clustering demonstrated differences in gene expression profiles of each metabolic disorder cohort versus CTRL and potential overlap amongst the signatures of the metabolic states, we further analyzed the relationships of gene expression within and amongst the 4 disease or pre-disease states in the context of gene sets. A gene set is defined as a group of genes with a common purpose, derived from the Gene Ontology project[31]. For further information on gene sets, normalization, and calculations, see the Methods section. Complete analysis with *p*-values for each gene set as well as the *p*-value and fold change for individual genes considered in each gene set comparison are also available (Supplemental Tables 1 and 2). Gene set analysis showed that genes driving the differential expression in MetS, CAD and T2D are associated with overlapping activation of the innate immune response, activation of the pro-inflammatory transcription factor NF-κB in CAD, and over-expression of genes involved in T cell activation and signaling in T2D.

Rheumatoid Arthritis

Rheumatoid Arthritis is an autoimmune disease characterized by systemic inflammation that extends into and damages peripheral joints[32]. Patients with RA have robust and distinguishable gene expression profiles in peripheral whole blood[22]. This finding was repeated using the HEEBO slide as the array format. Our analysis identified 5 gene sets of particular significance (Table 2). *BIRC4* is over-expressed in gene set 110, Cell Development, and is involved in activation of the transcription factor NF-κB. NF-κB regulates expression of many pro- inflammatory genes. Immune System Process, gene set 271, includes over-expression of *LAT2* and *NFAM1*, genes involved in B cell signaling and development. Additional genes, *BAT1, LIG4* and *ILF2*, are expressed in lymphocytes and differentially expressed in gene set 435. *BAT1* is an HLA-associated transcript mutated in patients with RA. *LIG4* encodes a protein essential for V(D)J recombination and non-homologous end joining as part of DNA repair. *ILF2* is involved in T cell expression of IL-2, a potent stimulator of proliferation of lymphocytes. The IL2-receptor alpha, *IL2RA*, is also over-expressed in this cohort and is found in gene set 753, Signal Transduction. Differential expression of genes involved in activation, maturation and signaling of lymphocytes is in agreement with the gene expression profile of RA seen previously[22]. Other genes significantly differentially expressed in this gene set included the IL9-receptor, *IL9R*, which supports IL-2 and IL-4 independent T cell growth, and *MAP2K6*, which activates p38 MAP kinase in response to inflammatory cytokines. *LILRB4* was significantly under-expressed as part of the Signal Transduction gene set. This gene is an immune-cell

receptor for MHC-I that transduces a signal to inhibit the immune response; increased expression of LILRB4 on antigen presenting cells renders the cells tolerant, therefore decreased expression might allow for increased autoreactivity[33]. Finally, in gene set 706, Response to External Stimulus, *CHST2*, encoding a protein expressed by vascular endothelium to attract lymphocytes, and *F11R*, encoding another protein expressed by vascular endothelium and involved in leukocyte transmigration, were over-expressed. Over-expression of genes encoding proteins with key roles in lymphocyte activation and growth could influence activation and expansion of self-reactive lymphocytes believed to cause joint destruction in individuals with RA.

Metabolic Syndrome

The triad of MetS, CAD, and T2D are typically considered metabolic, not immune, diseases although aspects of each involve inflammation, sometimes systemically. Nevertheless, each of these pathogenic states was characterized by an identifiable peripheral blood gene expression distinguishing each state from CTRL (Fig. 1). Here, we identify the differentially expressed genes driving these signatures.

The gene expression profile characterizing MetS consisted of many genes involved in innate immune responses (Table 2). Gene set 13, Acute Inflammatory Response, featured up-regulation of *CFHR1*, a complement factor gene, and *ORM1*, an acute phase reactant. Acute phase reactants may be present at increased levels as a consequence of hyperlipidemia-induced liver injury. *CD1D*, involved in antigen presentation of lipids and glycolipids to activate NKT cells, is over-expressed in gene set 316. Under-expression of *TNFAIP3* is found in gene set 407, Negative Regulation of Signal Transduction. This gene encodes a protein that inhibits NF-κB activation and terminates NF-κB responses. Decreased expression of this gene, as with *LILRB4* in RA, limits at least one way in which an immune response is attenuated. Also decreased in expression were two apoptosis-related genes: *DAXX*, involved in TNF-mediated apoptosis, and *MOAP1*, involved in caspase-mediated apoptosis, in the Regulation of Developmental Process gene set. *MAP3K5*, also in gene set 615, shows increased expression. *MAP3K5* activates *MAP2K6* which in turn activates p38 in response to inflammatory cytokines. This pathway was also over-expressed in RA.

Coronary Artery Disease

Peripheral blood gene expression in CAD was also distinguishable from the CTRL cohort (Fig. 1). This profile is defined by genes that impact activation and expression of NF-κB (Table 2). While some genes encoding proteins that impact NF-κB were differentially expressed in RA and MetS, the CAD gene expression profile encompassed a far greater number of NF-κB associated genes. Gene set 499, Positive Regulation of Immune Response, includes the over-expressed genes *IKBKG* and *TLR8*. *IKBKG* is a regulator of the IKK complex, which activates NF-κB; *TLR8* also activates NF-κB as part of the innate immune response. *MAP3K7IP2, TNFAIP3* and *TNFRSF10B* are differentially expressed in gene set 636. IL-1 initiated activation of NF-κB is mediated by *MAP3K7IP2, TNFRSF10B* is also an activator while previously mentioned *TNFAIP3,* an inhibitor of NF-κB, is under-expressed in this disease cohort, as well as in MetS. *TRIB3* in gene set 412, Negative Regulation of Transferase Activity, was highly over-expressed. This gene encodes a protein that is induced

by NF-κB and acts as a feedback regulator of this transcription factor, thus sensitizing the cells to apoptosis. Downstream effects of activation of NF-κB include increased expression of many genes involved in inflammation and also of genes that protect the immune cells from apoptosis, allowing further expansion of the inflammatory response.

Type 2 Diabetes

The T2D peripheral blood gene expression signature was robust and included many protein-coding genes involved in T cell signaling and function (Table 2). Three of these, gene sets 271, 435 and 753, were also significantly differentially expressed in RA and gene sets 435 and 753 specifically showcase the T cell associated genes. Gene set 435, Nucleobase Nucleoside Nucleotide and Nucleic Acid Metabolic Process, includes over-expression of the T cell genes *ILF2*, or nuclear factor of activated T cells, which modulates IL-2 expression, *NFATC4*, a gene involved in the inducible expression of cytokines and *NP*, a gene encoding an enzyme that, when lacking, compromises cell-mediated immunity. In gene set 753, Signal Transduction, the trio of receptors *IL1RL1, IL4R* and *IL9R* were over-expressed. *IL1RL1* is a receptor induced by inflammatory cytokines, *IL4R* promotes differentiation of T cells to T helper type 2 cells, and *IL9R* encodes a receptor that supports IL-2 and IL-4 independent growth of the T cell population. This gene set also features decreased expression of the leukocyte immunoglobulin-like receptors *LILRB2* and *LILRB4*, both of which serve to limit the immune response. *MAPK11,* encoding a protein activated by pro-inflammatory cytokines, is over-expressed in this gene set along with *GPX1*, a glutathione peroxidase. Finally, levels of *TNFRSF13B* transcripts are increased; this gene serves to stimulate lymphocyte function. A number of other gene sets were significantly differentially expressed in T2D. Cell Cell Signaling, gene set 104, includes up-regulation of the complement component *C1QA* and the chemotaxin *CXCL5*. Gene set 117 features increased expression of *CD276*, another regulator of T cell mediated immunity and *IL2RA*, a gene also over-expressed in RA and involved in proliferation of lymphocytes. The Immune System Process gene set 271 includes many of the previously discussed differentially regulated genes as well as decreased expression of *CTLA4,* a gene encoding a protein expressed on the surface of helper T cells that transduces an inhibitory signal. The gene expression profile of T2D was distinct from CTRL subjects in the differential expression of many genes involved in the activation of and signaling in T cells, reflecting the possibility that components of the adaptive immune system may contribute to the pathogenesis of T2D.

Correlation Among Disease States

To further investigate the overlap in gene expression profiles of the metabolic disorders suggested by hierarchical clustering (Fig. 2), we explored interrelationships of these profiles in the gene set analysis. To do so, we created a list of gene sets whose average expression level differed significantly from that of CTRL ($p<0.05$) for any of the four comparisons. Next, we assessed the relationships among RA, MetS, CAD and T2D by estimating pairwise Spearman Correlation coefficients based on the *p*-values for the gene sets derived from the comparison of each state to CTRL. The thickness of the line connecting one state to another is based on the estimated Spearman's correlations (Fig. 3A). The sole autoimmune disease, RA, showed the lowest correlation with the other diseases. We found the highest degrees of correlation in comparisons among MetS, CAD and T2D demonstrating strong overlap in the

peripheral blood gene expression profiles of these inflammatory disease states. Correlation among this trio ranged from Spearman's rho 0.44296 to 0.53772, all with significance of *p*<0.0001. There were 618 genes significantly differentially expressed in 2 or more of MetS, CAD or T2D versus CTRL comparisons (Fig. 3B). Within the genes differentially expressed in all three states versus CTRL, *FCGR1A*, an Fc receptor for immunoglobulin-gamma involved in both innate and adaptive immunity, *AGER*, a receptor for the immunogenic advanced glycation end products, the innate immunity-related complement stabilizer *CFP*, and the acute phase reactant, *CP*, were over-expressed. These genes and their related pathways may all lead to activation of the innate immune response. *PPARA*, a peroxisome proliferator receptor, also showed increased expression. *IL2RA*, the NF-κB activator *TNFRSF1A*, and the inflammatory signaling molecule *MAPK11* showed increased expression in both CAD and T2D. IL-1 mediates synthesis of acute phase reactants and the IL-1 receptor associated protein, *IL1RAP*, was differentially expressed in both MetS and T2D along with the NF-κB associated *NFKB2*. Differentially expressed in both CAD and MetS were the innate immune activator *LILRA5, MAP3K5* involved in the activation of p38 MAP kinase in response to inflammatory cytokines, and the NF-κB associated *NFKBIB*. Gene expression profiles of MetS, CAD and T2D were significantly correlated with each other and, to a lesser degree, with RA.

Given that MetS is a precursor to both CAD and T2D, an analysis was performed to eliminate those gene sets that overlap amongst CAD or T2D and MetS in order to isolate the genes and gene sets that may be involved in progression of MetS to its sequelae. We selected gene sets that were not significantly differentially expressed in MetS but were significantly differentially expressed in CAD or T2D (Table 2). As MetS progresses to its sequelae, CAD, we found differential expression of an increased number of genes involved in activation of and signaling in macrophages. The predominance of genes participating in activation of NF-κB, seen in the comparison of CAD to CTRL, was also found in the comparison of CAD to MetS. The increased expression of monocyte and macrophage related genes can be found primarily in gene set 753, Signal Transduction. *CD14* is a monocyte surface marker, *CXCL14* encodes a chemokine for monocytes, and *MST1R* encodes a protein that serves as the receptor for macrophage stimulating protein. All 3 of these genes were over-expressed in CAD compared to MetS. In addition, three MAP kinases, *MAP2K7, MAPK11* and *MAPK13*, were over-expressed in this gene set, all of which are involved in mediating the immune response to pro-inflammatory cytokines. Gene sets 482 and 682 feature a number of genes involved in the activation of the pro-inflammatory transcription factor NF-κB. *CARD14* interacts with *BCL10* to positively influence NF-κB activation; *IKBKG* and *TNFRSF1A* also activate NF-κB. Gene set 372, Negative Regulation of Biological Processes, contains differentially expressed *CLCF1*, a B cell stimulatory cytokine, *F2*, or coagulation factor II, associated with vascular inflammation, and *MPO,* encoding the protein myeloperoxidase, an enzyme found in neutrophils. In addition to the over-expression of NF-κB activating genes, also seen in the direct comparisons of CAD to CTRL and MetS, monocyte and macrophage related genes were also over-expressed in CAD.

Differences in peripheral blood gene expression of T2D as contrasted with MetS were much more subtle than the comparison of CAD with MetS. Of note are *CD276, LAT* and *LCK*, in gene sets 271, 478 and 596. *CD276* is involved in regulation of cell-mediated immune responses in T cells, *LAT* is a component of the cell surface T cell receptor complex, and *LCK* is a protein involved in the maturation of T cells. Also significant, *ILF2*, encoding a protein that regulates IL-2 and proliferation of T cells, was over-expressed in T2D relative to MetS. Finally, *CXCL5*, a chemotactic cytokine, was increased in expression in gene set 104, Cell Cell Signaling. T2D and MetS were the two most closely correlated disease states (Fig. 3A). The gene expression profiles of these two states differ primarily in the over-expression of T cell associated genes in T2D.

PCR Validation

To quantitatively measure differences in transcript levels of a select group of genes identified by the array analysis, we performed quantitative-reverse transcriptase PCR (RT-PCR). We analyzed 19 of the original 35 samples used for the microarray analysis (group 1). In addition, we obtained 61 independent samples from CTRL, MetS, CAD and T2D subjects (group 2). We determined the fold difference between each experimental group and its own CTRL group, eg. group 1 or group 2, using the   Ct method (Table 3). A 'pooled' p value was calculated by pooling results from groups 1 and 2. From the MetS peripheral blood gene expression profile, *CD1D* also showed increased expression while the decreased expression of *DAXX, MOAP1* and *TNFAIP3* was similarly validated. Of interest, our additional analysis demonstrates that *CD1D* and *DAXX* were also differentially expressed in the CAD cohort. Expression of *MOAP1* and *TNFAIP3* was also decreased in all three metabolic cohorts relative to the level in the CTRL cohort in both group1 and group 2. Three genes over-expressed in the CAD signature were also confirmed by RT-PCR measurements, *CD14, CFHR1* and *CXCL14*. These genes also showed significant differential expression in MetS (*CFHR1*), T2D (*CXCL14*) or both (*CD14*). Finally, the differences in expression of *CTLA4, GPX1, IL4R* and *NP*, from the T2D microarray signature, were confirmed by the RT-PCR experiments. *CTLA4* also displayed decreased transcript levels in CAD and *GPX1* and *IL4R* showed increased and decreased expression, respectively, in all 3 metabolic cohorts. Besides validating results obtained from microarray analyses in independent cohorts by an independent method, these experiments also identify expression patterns of individual genes unique to one or two metabolic disorders or shared by all three metabolic disorders.

Discussion

Our analysis of peripheral blood gene expression in CAD, T2D and their precursor state, MetS, by microarray shows that these disorders feature unique gene expression signatures. We included individuals with RA in these studies as an example of a disease with a known peripheral blood gene expression profile and for purposes of comparing the metabolic expression signatures to that of an autoimmune disease. As expected, gene expression of the RA cohort was sufficient to distinguish these individuals from CTRL. In each of MetS, CAD and T2D, there were sufficient numbers of genes differentially expressed to cluster the majority of each group away from the CTRL cohort with 100% support. Additionally, when all 4 disease states were included in the analysis, 24/26 subjects from the disease cohorts

branched together with 100% support. Within that branch, the RA patients clustered together while the metabolic cohorts showed considerable overlap. Thus, the metabolic cohorts have peripheral blood gene expression signatures that are more similar to RA than CTRL, but also more similar to each other than RA.

The gene expression signature of MetS centers on dysregulation of genes involved in the innate immune response. One component of MetS is hypercholesterolemia, specifically, greater levels of very low density lipoprotein (VLDL). VLDL stimulates release of acute phase proteins from the liver. Activation of the innate immune response in peripheral blood could be a response to increased amounts of circulating VLDL. Fatty acids are known to activate innate immune signaling molecules, like TLR4[34]. The gene expression signature of MetS shares much in common with that of CAD and T2D; many of the gene sets differentially expressed in the individual comparisons of MetS, CAD and T2D to CTRL also overlap among the three disorders. Spearman's test for correlation showed clear association of the three metabolic disorders, an association that was also significant, but to a lesser extent when correlated to RA. The gene sets and corresponding genes driving this similarity are those associated with activation of the innate immune response, an association not seen in the RA cohort.

In addition to activation of the innate immune response, many genes involved in activation of the pro-inflammatory transcription factor, NF-κB, are differentially expressed in the CAD profile. Comparing CAD and T2D directly to their precursor, MetS, is a more appropriate analysis to determine genes and pathways involved in progression of pre-disease to disease. The comparison of CAD to MetS revealed that monocyte and macrophage associated genes are more prominently differentially expressed. In addition to the hyperlipidemia of MetS, diagnosis of CAD indicates the presence of atherosclerotic plaques in the lumen of peripheral blood vessels. CAD gene expression profiles uncovered here reflect systemic inflammation and activation of monocytes. Many of these activated monocytes may migrate from the lumen to become the lipid-filled macrophages seen in the core of these plaques[35]. One possible interpretation of these results is that immunological processes occurring at the site of disease are reflected in peripheral blood.

In the gene expression profile of T2D, a disease that represents more refractory insulin resistance than MetS, we see increased expression of genes associated with activation, signaling and function of T cells. This was also the case in a direct comparison of gene expression between MetS and T2D. Many of these T cell activation genes are also differentially expressed in RA; however unlike T2D, in RA there is a documented role of T cells in the pathogenesis of disease: as the effector cells of joint-specific destruction[32]. The up-regulation of T cell activation seen in these studies may be a byproduct of enhanced activation of the immune response by adipocytes. Recent studies have shown activated T cells to be present in abundance in visceral adipose tissue of mice with T2D[36].

This independent study also replicates a number of findings in the literature with regards to altered expression of genes in states of insulin resistance and obesity. The monocyte surface antigen CD14, upregulated in MetS, CAD and T2D is also upregulated in mice with insulin resistance[37]. CXCL14 null female mice are protected from obesity-induced hyperglycemia

and do not develop insulin resistance[38]. Our results show that expression of *CXCL14* is elevated in both CAD and T2D relative to CTRL raising the possibility that elevated CXCL14 may also contribute to insulin resistance in human disease. Additional correlations can be seen in human studies showing that a SNP in the *IL4R* gene is associated with increased body mass index[39] an our studies documenting decreased *IL4R* expression in MetS, CAD and T2D.

Taken together, our data support a hypothesis whereby MetS produces a state of general systemic inflammation mediated by the innate immune system. This inflammation persists as the pre-disease state progresses to CAD or T2D. Peripheral blood gene expression in CAD and T2D identifies additional immune processes that may underlie these two disease phenotypes; NF-κB activation in CAD, T cell activation in T2D; however, PCR analysis of these genes in independent cohorts suggest that differential expression may overlap in MetS, CAD and T2D to a greater extent than was evident by microarray analysis. This result is in accordance with the overlap in biological processes and risk factors underlying the disease states. However, in spite of this, there was enough uniqueness to separate the cohorts by microarray. PCR analysis allowed us to further investigate these differences. A certain relatedness could be seen amongst MetS, CAD and T2D, for example, the expression of CXCL14 appears to be on a gradient where expression was increased around 3-fold in group 1 MetS, and greater than 10-fold in CAD and T2D (Table 3). Another reason for this may be the overlap in criteria defining each disease and propensity for having undiagnosed CAD in a patient with T2D, for example. While the advanced statistics in the microarray analyses detected differences enough to delineate MetS, CAD and T2D, the overlap in criteria and incidence of these three states was more accurately defined in the quantitative PCR experiments. PCR validated many of our initial findings and thus, the gene expression profiles of MetS, CAD and T2D present convincing evidence that systemic inflammation is a component of the pathogenesis of all 3 states. Furthermore, this study identifies a minimally invasive system that could be used in longitudinal studies to better understand progression of MetS to its sequelae.

Materials and Methods

Patient Recruitment

Rheumatoid arthritis is defined by the American College of Rheumatology Criteria. Patients displayed four or more of the following symptoms for greater than 6 months: morning stiffness, swelling in 3 or more joints, swelling of finger and/or wrist joints, symmetric swelling, rheumatoid nodules, positive rheumatoid factor, or radiographic erosions in the hand and/or wrist[40]. *Metabolic syndrome* is defined by the International Federation of Diabetes as central obesity plus any 2 of the following 4 characteristics: hypertriglyceridemia, low HDL cholesterol, hypertension or raised fasting plasma glucose[41]. C*oronary artery disease* was diagnosed in each patient using imaging techniques to detect flow-limiting coronary artery stenoses[42]. Three of the 6 patients with coronary artery disease participating in this study were post coronary artery bypass graft or myocardial infarction. All patients in this cohort are also being treated for systemic hypertension. *Diabetes* is defined by the WHO criteria of classic symptoms of diabetes (polydipsia, polyuria,

polyphagia and weight loss) and a plasma glucose >200 mg/dl, a fasting plasma glucose of >126 mg/dl or a 2 h plasma glucose during an oral glucose tolerance test of >200 mg/dl. Type 1 diabetes is differentiated from T2D by a number of clinical criteria including history, clinical presentation and laboratory findings. Type 2 diabetics are more likely to have a high body mass index and less likely to need insulin in restoring normal plasma glucose levels[5]. *Control* patients have not ever received any of the previous diagnoses, have not been diagnosed with any autoimmune or other chronic disease, and are not currently taking medication for any illness or condition. The study was approved by the Institutional Review Board of Vanderbilt University and all subjects provided written informed consent.

Microarray Gene Expression Experiments

Peripheral whole blood was drawn directly into PreAnalytiX PAXgene tubes (VWR, West Chester, PA). RNA was isolated using the PreAnalytiX protocol "Manual Purification of Total RNA from Human Whole Blood Collected into PAXgene Blood RNA Tubes." Amplified CTRL and sample RNA was coupled to Cy3 or Cy5 dyes (GE Healthcare, Piscataway, NJ), respectively, using the Vanderbilt Functional Genomics Shared Resource (FGSR) coupling protocol, found at [array.mc.vanderbilt.edu]. The reverse transcription reaction used 6 μg of Oligo dT and the superscript III reverse transcriptase (Invitrogen, Carlsbad, CA). Labeled cDNA was purified using the Qiagen QiaQuick PCR purification kit and resuspended in 2X hybridization buffer (50% formamide, 10X SSC and 0.2% SDS) and 1 μl polyA RNA. Labeled, resuspended cDNA was heated to 100°C for 2 min and hybridized to the Human Exonic Evidence-Based Oligonucleotide (HEEBO) array at 42°C for 16 h in a heating oven. The HEEBO slide was designed by the Stanford Functional Genomics Facility (microarray.org). Oligo probes are commercially available (Invitrogen, Carlsbad, CA) and the slides were printed by Microarrays Inc (Hudson Alpha Institute, Huntsville, AL). We washed and dried the slides per the FGSR protocol and scanned them into the GenePix Pro4.1 Software using a 400B scanner (Axon Instruments, Union City, CA). We analyzed intensity data using GenePix software in combination with The Institute for Genomic Research's TM4: Microarray Suite programs[43].

Microarray Data Analysis

The Institute for Genomic Research's Multi-Experiment Viewer was used to visualize intensity data. We used Significance Analysis of Microarray (SAM) to determine a group of significantly under- and over-expressed genes in the comparisons of each disease group versus CTRL. Each group's input consisted of array intensity values normalized to an average intensity of 0.20. For each comparison the individual disease cohort (Fig. 1) or all disease cohorts combined (Fig. 2) were assigned to group A while the control cohort was assigned to group B and the SAM analysis was performed with 1,000 permutations. The median number of falsely significant genes was set to   2. The results of these analyses served as the input for support tree analysis using the bootstrap statistical method with an additional 1,000 permutations, which created hierarchical clustering trees for the 4 comparisons of each disease versus CTRL as well as for the comparison of all diseases versus control. . For statistical analysis of gene sets, we normalized microarray data using the print-tip lowess normalization algorithm as implemented in the Bioconductor package *marray*[44]. We used maximum expression levels from multiple probe sets corresponding to

the same gene to represent the gene expression level. To ensure reliable gene expression estimates, we included genes with intensity values for more than 6 CTRL samples and more than 3 samples for each of the other groups. There were 14,558 genes left after this step. To identify groups of functionally related genes differentially expressed for different patient groups, we conducted gene set analysis using the mixed effects models approach[45; 46]. Gene sets used in these analyses were derived from the controlled vocabulary of the Gene Ontology (GO) project, http://www.broad.mit.edu/gsea/msigdb/index.jsp. For each gene set, the mixed models included gene expression levels as outcome, group (disease group vs. CTRL group) as the fixed effect and batches as the random effects. In addition, we included random effects based on eigenvectors of gene-gene correlation matrix to account for correlation patterns of the genes[46]. Because we examined many gene sets, to control for the rate of false positive findings by chance, we adjusted nominal *p*-values using the method of false discovery rate[47]. To study the relations between T2D, MetS, CAD and RA, we estimated pairwise Spearman correlation coefficients for these disease groups based on nominal pathway *p*-values from comparing each disease group versus CTRL. We used Cytoscape software[48] to visualize these associations.

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus[49] and are accessible through GEO Series accession number GSE23561 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE23561).

RT-PCR

RT-PCR was performed on a subset of samples from the original microarray analysis and a subset of new samples. Samples were selected from the original experiments for further testing based on the sole criteria of sufficient RNA remaining to perform RT-PCR analysis. All samples with sufficient RNA were analyzed by RT-PCR. Gene expression was determined by RT-PCR using a TaqMan Low Density Array (TLDA). Fold change expression levels were determined by the Ct method, comparing expression of test gene to an average of two independent measurements of GAPDH, and then comparing the disease cohort versus CTRL. Significance was determined using a t-test on the Ct raw values.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Stumvoll M, Goldstein BJ, van Haeften TW. Type 2 diabetes: principles of pathogenesis and therapy. Lancet. 2005; 365(9467):1333–1346. [PubMed: 15823385]

2. National diabetes fact sheet: general information and national estimates on diabetes in the United States, 2007. US Department of Health and Human Services, Centers for Disease Control and Prevention; Atlanta, GA: 2008.

3. Ross R. The pathogenesis of atherosclerosis: a perspective for the 1990s. Nature. 1993; 362(6423): 801–809. [PubMed: 8479518]

4. Lloyd-Jones D, Adams R, Carnethon M, De Simone G, Ferguson TB, Flegal K, et al. Heart Disease and Stroke Statistics--2009 Update: A Report From the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation. 2009; 119(3):e21–181. [PubMed: 19075105]

5. Alberti KGMM, Zimmet P, Shaw J. The metabolic syndrome--a new worldwide definition. Lancet. 2005; 366(9491):1059–1062. [PubMed: 16182882]

6. Ford ES. Prevalence of the Metabolic Syndrome Defined by the International Diabetes Federation Among Adults in the U.S. Diabetes Care. 2005; 28(11):2745–2749. [PubMed: 16249550]

7. Galassi A, Reynolds K, He J. Metabolic Syndrome and Risk of Cardiovascular Disease: A Meta-Analysis. Am J Med. 2006; 119(10):812–819. [PubMed: 17000207]

8. Gami AS, Witt BJ, Howard DE, Erwin PJ, Gami LA, Somers VK, et al. Metabolic Syndrome and Risk of Incident Cardiovascular Events and Death: A Systematic Review and Meta-Analysis of Longitudinal Studies. J Am Coll Cardiol. 2007; 49(4):403–414. [PubMed: 17258085]

9. Ford ES, Li C, Sattar N. Metabolic Syndrome and Incident Diabetes. Diabetes Care. 2008; 31(9): 1898–1904. [PubMed: 18591398]

10. Sattar N, McConnachie A, Shaper AG, Blauw GJ, Buckley BM, de Craen AJ, et al. Can metabolic syndrome usefully predict cardiovascular disease and diabetes? Outcome data from two prospective studies. Lancet. 2008; 371(9628):1927–1935. [PubMed: 18501419]

11. Meigs JB, Wilson PWF, Fox CS, Vasan RS, Nathan DM, Sullivan LM, et al. Body Mass Index, Metabolic Syndrome, and Risk of Type 2 Diabetes or Cardiovascular Disease. J Clin Endocrinol Metab. 2006; 91(8):2906–2912. [PubMed: 16735483]

12. Shimabukuro M. Cardiac Adiposity and Global Cardiometabolic Risk New Concept and Clinical Implication. Circ J. 2009; 73(1):27–34. [PubMed: 19057089]

13. Zimmet P, Alberti KGMM, Shaw J. Global and societal implications of the diabetes epidemic. Nature. 2001; 414(6865):782–787. [PubMed: 11742409]

14. Wellen KE, Hotamisligil GS. Obesity-induced inflammatory changes in adipose tissue. J Clin Invest. 2003; 112(12):1785–1788. [PubMed: 14679172]

15. Martinez FO, Gordon S, Locati M, Mantovani A. Transcriptional Profiling of the Human Monocyte-to-Macrophage Differentiation and Polarization: New Molecules and Patterns of Gene Expression. J Immunol. 2006; 177(10):7303–7311. [PubMed: 17082649]

16. Tan Q, Zhao J, Li S, Christiansen L, Kruse TA, Christensen K. Differential and correlation analyses of microarray gene expression data in the CEPH Utah families. Genomics. 2008; 92(2): 94–100. [PubMed: 18519161]

17. Gregg JP, Lit L, Baron CA, Hertz-Picciotto I, Walker W, Davis RA, et al. Gene expression changes in children with autism. Genomics. 2008; 91(1):22–29. [PubMed: 18006270]

18. Lockstone HE, Harris LW, Swatton JE, Wayland MT, Holland AJ, Bahn S. Gene expression profiling in the adult Down syndrome brain. Genomics. 2007; 90(6):647–660. [PubMed: 17950572]

19. Sotiriou C, Pusztai L. Gene-Expression Signatures in Breast Cancer. New England Journal of Medicine. 2009; 360(8):790–800. [PubMed: 19228622]

20. Cheok MH, Yang W, Pui C-H, Downing JR, Cheng C, Naeve CW, et al. Treatment-specific changes in gene expression discriminate in vivo drug response in human leukemia cells. Nat Genet. 2003; 34(1):85–90. [PubMed: 12704389]

21. Holleman A, Cheok MH, den Boer ML, Yang W, Veerman AJP, Kazemier KM, et al. Gene-Expression Patterns in Drug-Resistant Acute Lymphoblastic Leukemia Cells and Response to Treatment. N Engl J Med. 2004; 351(6):533–542. [PubMed: 15295046]

22. Aune TM, Maas K, Moore JH, Olsen NJ. Gene expression profiles in human autoimmune disease. Current pharmaceutical design. 2003; 9(23):1905–1917. [PubMed: 12871194]

23. 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. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(5):2610–2615. [PubMed: 12604793]

24. Bomprezzi R, Ringner M, Kim S, Bittner ML, Khan J, Chen Y, et al. Gene expression profile in multiple sclerosis patients and healthy controls: identifying pathways relevant to disease. Hum Mol Genet. 2003; 12(17):2191–2199. [PubMed: 12915464]

25. Maas K, Chen H, Shyr Y, Olsen NJ, Aune T. Shared gene expression profiles in individuals with autoimmune disease and unaffected first-degree relatives of individuals with autoimmune disease. Human molecular genetics. 2005; 14(10):1305–1314. [PubMed: 15814587]

26. Liu Z, Maas K, Aune TM. Identification of gene expression signatures in autoimmune disease without the influence of familial resemblance. Human molecular genetics. 2006; 15(3):501–509. [PubMed: 16371420]

27. Maas K, Westfall M, Pietenpol J, Olsen NJ, Aune T. Reduced p53 in peripheral blood mononuclear cells from patients with rheumatoid arthritis is associated with loss of radiation-induced apoptosis. Arthritis and rheumatism. 2005; 52(4):1047–1057. [PubMed: 15818671]

28. Heap G, Trynka G, Jansen R, Bruinenberg M, Swertz M, Dinesen L, et al. Complex nature of SNP genotype effects on gene expression in primary human leucocytes. BMC medical genomics. 2009; 2(1):1. [PubMed: 19128478]

29. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009; 106(23):9362–9367. [PubMed: 19474294]

30. Torkamani A, Topol EJ, Schork NJ. Pathway analysis of seven common diseases assessed by genome-wide association. Genomics. 2008; 92(5):265–272. [PubMed: 18722519]

31. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000; 25(1):25–29. [PubMed: 10802651]

32. Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. Lancet. 2009; 373(9664):659–672. [PubMed: 19157532]

33. Chang CC, Ciubotariu R, Manavalan JS, Yuan J, Colovai AI, Piazza F, et al. Tolerization of dendritic cells by TS cells: the crucial role of inhibitory receptors ILT3 and ILT4. Nat Immunol. 2002; 3(3):237–243. [PubMed: 11875462]

34. Shi H, Kokoeva MV, Inouye K, Tzameli I, Yin H, Flier JS. TLR4 links innate immunity and fatty acid-induced insulin resistance. J Clin Invest. 2006; 116(11):3015–3025. [PubMed: 17053832]

35. Berliner JA, Navab M, Fogelman AM, Frank JS, Demer LL, Edwards PA, et al. Atherosclerosis: Basic Mechanisms : Oxidation, Inflammation, and Genetics. Circulation. 1995; 91(9):2488–2496. [PubMed: 7729036]

36. Nishimura S, Manabe I, Nagasaki M, Eto K, Yamashita H, Ohsugi M, et al. CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. Nat Med. 2009; 15(8):914–920. [PubMed: 19633658]

37. Yessoufou A, Moutairou K, Khan NA. A Model of Insulin Resistance in Mice, Born to Diabetic Pregnancy, Is Associated with Alterations of Transcription-Related Genes in Pancreas and Epididymal Adipose Tissue. E-pub 2010 Sep 26.

38. Hara, T.; Nakayama, Y.; Gerald, L. Vitamins & Hormones. Vol. 80. Academic Press; 2009. Chapter 5 CXCL14 and Insulin Action; p. 107-123.

39. Ha E, Yang S-H, Yoo K-I, Chung I-S, Lee M-Y, Bae J-H, et al. Interleukin 4 receptor is associated with an increase in body mass index in Koreans. Life Sciences. 2008; 82(19-20):1040–1043. [PubMed: 18433792]

40. Arnett FC, Edworthy SM, Bloch DA, Mcshane DJ, Fries JF, Cooper NS, et al. The american rheumatism association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988; 31(3):315–324. [PubMed: 3358796]

41. Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. Lancet. 2005; 365(9468):1415–1428. [PubMed: 15836891]

42. Min JK, Shaw LJ. Noninvasive Diagnostic and Prognostic Assessment of Individuals With Suspected Coronary Artery Disease: Coronary Computed Tomographic Angiography Perspective. Circ Cardiovasc Imaging. 2008; 1(3):270–281. [PubMed: 19808551]

43. Saeed A, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. Biotechniques. 2003; 34(2):374–378. [PubMed: 12613259]

44. Reimers, M.; Carey, VJ.; Alan, K.; Brian, O. Methods in Enzymology. Vol. 411. Academic Press; 2006. Bioconductor: An Open Source Framework for Bioinformatics and Computational Biology; p. 119-134.

45. Wang L, Zhang B, Wolfinger RD, Chen X. An Integrated Approach for the Analysis of Biological Pathways using Mixed Models. PLoS Genet. 2008; 4(7):e1000115. [PubMed: 18852846]

46. Wang L, Chen X, Wolfinger RD, Franklin JL, Coffey RJ, Zhang B. A Unified Mixed Effects Model for Gene Set Analysis of Time Course Microarray Experiments. Stat Appl Genet Mol Biol. 2009; 8(1) Article 47.

47. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol. 1995; 57(1):289–300.

48. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003; 13(11):2498–2504. [PubMed: 14597658]

49. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research. 2002; 30(1):207–210. [PubMed: 11752295]
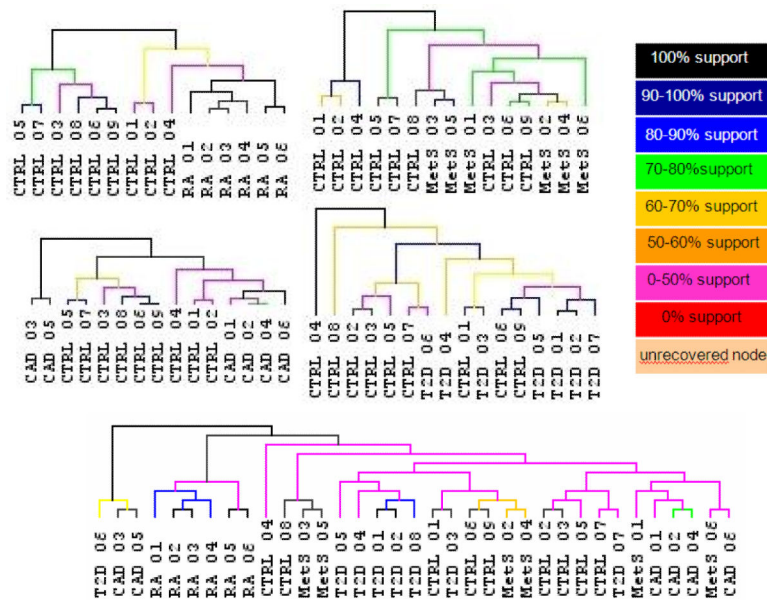
Fig. 1.

Unsupervised hierarchical clustering of individual disease cohorts with CTRL. .To determine if differential patterns of gene expression could be found among combinations of samples, normalized intensity data points from oligos with an average intensity of 0.20 (average array intensity) were inputted into The Institute for Genomic Research's Multi-Experiment Viewer. For each comparison, gene intensity averages were calulcated and those 0.20 were selected as input in each comparison. The CTRL v RA input was 4,969 gene and gene splice data points; CTRL v MetS input was 4,225 data points; CTRL v CAD input contained 4,271 data points and the CTRL v T2D comparison featured an input of 4,983 data points. For the comparison of all disease cohorts and CTRL, the input was 40,538 data points. These samples were inputted into a bootstrap analysis resulting in the hierarchical clustering trees shown in this figure. Statistical support for each branch of the tree is shown by color, legend to the right. CTRL= control, RA= rheumatoid arthritis, T2D= type 2 diabetes, MetS= metabolic syndrome and CAD= coronary artery disease.
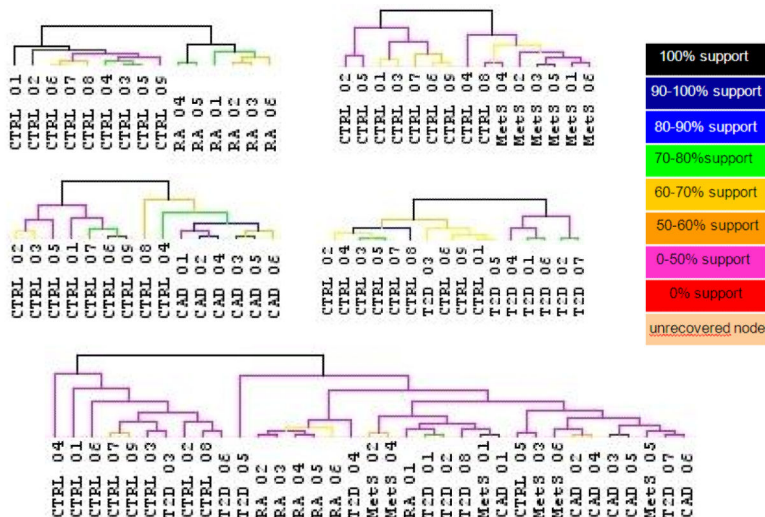
Fig. 2.
Supervised hierarchical clustering of cohorts versus CTRL. To determine if the gene
expression profiles of the disease cohorts were distinguishable from that of the 9 CTRL
patients and determine the similarity and difference of the profiles of each disease cohort to
each other in the presence of CTRL, the groups from Fig. 1 were analyzed by significance
analysis of microarray, with a median number of falsely significant genes set to   2. This
yielded lists of significant genes in each comparison, This list was inputted into a bootstrap
analysis resulting in the hierarchical clustering trees shown. CTRL= control, RA=
rheumatoid arthritis, T2D= type 2 diabetes, MetS= metabolic syndrome and CAD= coronary
artery disease.
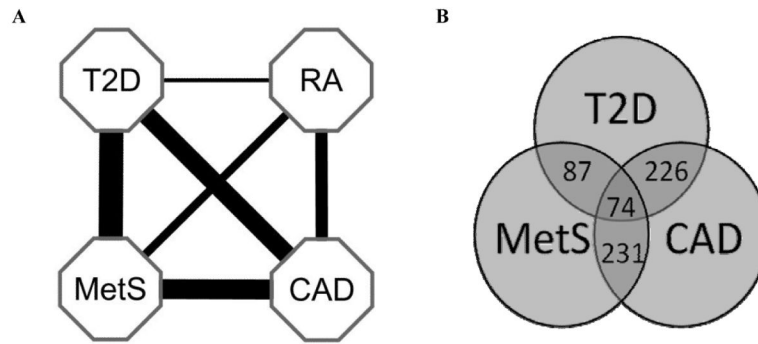
Fig. 3.
Correlative relationships among disease cohort gene expression. (A) Gene sets that significantly differed in expression versus CTRL were the input for this Spearman's correlation coefficient based diagram. Thickness of the bar represents a combination of Spearman's rho and statistical significance of the correlation. RA= rheumatoid arthritis, T2D= type 2 diabetes, MetS= metabolic syndrome and CAD= coronary artery disease. For the RA-T2D comparison Spearman's rho=0.10396, p=0.0555, RA-CAD rho=0.28462, p<0.0001, RA-MetS rho=0.19942, p=0.0002. T2D compared to CAD rho=0.42389, p<0.0001, T2D-MetS rho=0.53772, p<0.0001 and for the comparison of CAD to MetS rho=0.44296, p<0.0001. (B) A Venn diagram representing the number of genes with significantly different expression in each disease state versus CTRL that overlap among 2 or more of the states.

Table 1

SNPs associated with RA and T2D show differential gene expression

		RA		T2D	
RA SNP	**Gene**	**p^1**	**FC^2**	**p**	**FC**
Rs6682654	*CD244*	0.009	5.57	ns	
Rs2104286	*IL2RA*	0.002	4.30	0.028	2.52
$-^3$	*PRKCA*	0.001	2.88	ns	
-	*SLC22A4*	0.044	0.32	ns	
Rs3761847	*TRAF1*	0.026	0.34	0.010	0.29
T2D SNP					
Rs4607103	*ADAMTS9*	0.003	11.69	0.030	4.65
Rs2789686	*ANXA11*	ns		0.028	0.016
Rs2237892	*KCNQ1*	0.049	0.59	0.020	0.58

ns= not significant

[1] *p*= derived from Mixed Effects Model, RA or T2D relative to CTRL (ref. 39,40)

[2] FC= fold change, average of RA or T2D cohort relative to average of CTRL

[3] = identified via pathway-based analysis in Torkamani, *et al.*

Table 2

Differentially expressed gene sets

Gene Set	Gene Set Name	*p*-value
RA v CTRL		
110	Cell Development	0.0045
271	Immune System Process	0.0116
435	Nucleobase Nucleoside Nucleotide and Nucleic Acid Metabolic Process	2.13E-08
706	Response to External Stimulus	0.0078
753	Signal Transduction	1.03E-11
MetS v CTRL		
13	Acute Inflammatory Response	0.048
316	Lymphocyte Differentiation	0.004
407	Negative Regulation of Signal Transduction	0.051
615	Regulation of Developmental Process	0.023
CAD v CTRL		
412	Negative Regulation of Transferase Activity	0.014
499	Positive Regulation of Immune Response	0.020
636	Regulation of I KappaB Kinase NF KappaB Cascade	0.051
T2D v CTRL		
104	Cell Cell Signaling	0.0048
117	Cell Proliferation Go 0008283	0.002
271	Immune System Process	1.7E-06
435	Nucleobase Nucleoside Nucleotide and Nucleic Acid Metabolic Process	9.7E-28
753	Signal Transduction	4.8E-13
CAD v MetS		
372	Negative Regulation of Biological Process	5.7E-04
482	Positive Regulation of Cellular Process	0.008
682	Regulation of Transcription	0.019
753	Signal Transduction	0.009
T2D v MetS		
271	Immune System Process	0.043
478	Positive Regulation of Caspase Activity	0.033
596	Regulation of Cellular Metabolic Process	0.052
104	Cell Cell Signaling	0.030

Table 3

RT-PCR determined ratios[1] of differentially expressed genes

Gene	MetS			CAD			T2D		
	Group 1[2]	Group 2[3]	p-value[4]	Group 1	Group 2	p-value	Group 1	Group 2	p-value
CD14	1.88	1.33	**0.008**	1.68	1.26	**0.005**	1.39	1.39	**0.009**
CD1D	1.59	1.36	**0.006**	1.49	1.35	**0.003**	0.55	0.83	ns
CFHR1	9.21	3.85	**<0.0001**	3.48	3.08	**<0.0001**	1.34	1.48	ns
CTLA4	1.52	0.65	ns	0.21	0.41	**0.002**	0.55	0.58	**0.002**
CXCL14	3.4	0.32	ns	11.82	10.95	**0.002**	10.3	17.95	**0.002**
DAXX	0.33	0.43	**<0.0001**	0.53	0.51	**<0.0001**	0.88	0.72	ns
GPX1	4.75	1.57	**0.002**	2.41	1.61	**0.003**	1.91	2.05	**0.003**
IL4R	0.61	0.51	**<0.0001**	0.26	0.35	**<0.0001**	0.49	0.68	**0.008**
MOAP1	0.42	0.28	**<0.0001**	0.31	0.67	**0.01**	0.72	0.62	**0.03**
NP	2.31	1.06	ns	1.36	1.25	ns	1.82	1.41	**0.02**
TNFAIP3	0.4	0.26	**<0.0001**	0.23	0.27	**<0.0001**	0.97	0.41	**0.03**

ns= not significant

[1] ratio= fold change, determined by Ct calculations, calculated separately for each group versus group-specific CTRLs

[2] Group 1 is composed of 19 samples used in the original geneset analysis (CTRL=4, MetS=6, CAD=3, T2D=6)

[3] Group 2 is an independent set of 61 patient samples (CTRL=16, MetS=16, CAD=13, T2D=16)

[4] p-values calculated on groups 1 and 2 pooled data