
Copyright: © 2018 Macías-Segura et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Transcriptional signature associated with early rheumatoid arthritis and healthy individuals at high risk to develop the disease

N. Macías-Segura\textsuperscript{1,2*}, J. E. Castañeda-Delgado\textsuperscript{3}, Y. Bastian\textsuperscript{3}, D. Santiago-Algarra\textsuperscript{1,4}, J. D. Castillo-Ortiz\textsuperscript{5}, A. L. Alemán-Navarro\textsuperscript{1,2}, E. Jaime-Sánchez\textsuperscript{1}, M. Gomez-Moreno\textsuperscript{1,2}, C. A. Saucedo-Toral\textsuperscript{1}, Edgar E. Lara-Ramírez\textsuperscript{1}, M. Zapata-Zuñiga\textsuperscript{6}, L. Encisco-Moreno\textsuperscript{1}, R. González-Amaro\textsuperscript{5}, C. Ramos-Remus\textsuperscript{6,7}, J. A. Enciso-Moreno\textsuperscript{1*}

\textsuperscript{1} Unidad de Investigación Biomédica de Zacatecas, IMSS, Zacatecas, México, \textsuperscript{2} Departamento de Inmunología, Facultad de Medicina, Universidad Autónoma de San Luis Potosí, San Luis Potosí, México, \textsuperscript{3} Catedras-CONACYT, Unidad de Investigación Biomédica de Zacatecas-IMSS, Zacatecas, México, \textsuperscript{4} Instituto Potosino de Investigaciones Científicas y Tecnológicas, San Luis Potosí, México, \textsuperscript{5} Unidad de Investigación en Enfermedades Crónico-Degenerativas, Guadalajara, México, \textsuperscript{6} Facultad de Medicina Humana y Ciencias de la Salud, Universidad Autónoma de Zacatecas, Hospital Rural #51, IMSS, Villanueva, Zacatecas, México, \textsuperscript{7} Universidad Autónoma de Guadalajara, Guadalajara, México

\* Current address: Departamento de Fisiología y Farmacología, Centro de Ciencias Básicas, Universidad Autónoma de Aguascalientes, México

\* enciso_2000@yahoo.com

Abstract

Background

Little is known regarding the mechanisms underlying the loss of tolerance in the early and preclinical stages of autoimmune diseases. The aim of this work was to identify the transcriptional profile and signaling pathways associated to non-treated early rheumatoid arthritis (RA) and subjects at high risk. Several biomarker candidates for early RA are proposed.

Methods

Whole blood total RNA was obtained from non-treated early RA patients with <1 year of evolution as well as from healthy first-degree relatives of patients with RA (FDR) classified as ACCP+ and ACCP- according to their antibodies serum levels against cyclic citrullinated peptides. Complementary RNA (cRNA) was synthetized and hybridized to high-density microarrays. Data was analyzed in Genespring Software and functional categories were assigned to a specific transcriptome identified in subjects with RA and FDR ACCP positive. Specific signaling pathways for genes associated to RA were identified. Gene expression was evaluated by qPCR. Receiver operating characteristic (ROC) analysis was used to evaluate these genes as biomarkers.

Results

A characteristic transcriptome of 551 induced genes and 4,402 repressed genes were identified in early RA patients. Bioinformatics analysis of the data identified a specific transcriptome in RA patients. Moreover, some overlapped transcriptional profiles between patients
with RA and ACCP+ were identified, suggesting an up-regulated distinctive transcriptome from the preclinical stages up to progression to an early RA state. A total of 203 pathways have up-regulated genes that are shared between RA and ACCP+. Some of these genes show potential to be used as progression biomarkers for early RA with area under the curve of ROC > 0.92. These genes come from several functional categories associated to inflammation, Wnt signaling and type I interferon pathways.

Conclusion
The presence of a specific transcriptome in whole blood of RA patients suggests the activation of a specific inflammatory transcriptional signature in early RA development. The set of overexpressed genes in early RA patients that are shared with ACCP+ subjects but not with ACCP- subjects, can represent a transcriptional signature involved with the transition of a preclinical to a clinical RA stage. Some of these particular up-regulated and down-regulated genes are related to inflammatory processes and could be considered as biomarker candidates for disease progression in subjects at risk to develop RA.

Introduction
Rheumatoid arthritis (RA) has a multidimensional effect including progressive destruction of diarthrodial joints, development of comorbid conditions, family distress and high social costs [1–3]. Although the incidence of RA peaks during the fifth decade of life, in some countries close to 50% of patients develop symptoms before the age of 35 [4]. Early treatment can limit the overall impact of the disease [5, 6], including the prevention of joint damage and work absence [7–9]. Therefore, detection of RA at very early stages, or even during the pre-clinical phase of the disease, would have a meaningful impact on the patient outcome.

The genetic heritability of RA has been estimated at 12% to 60% [10]. Variation on these figures may be explained by differences in study design as well as in ethnicity and in characteristics such as duration of disease and treatments. A number of polymorphisms and variants in HLA molecules have been associated with the development of this disease, particularly the so-called “Shared epitope” [2, 4, 8]. Identification of a particular transcriptional signature in preclinical high-risk subjects would add insight on the molecular mechanisms involved in the development of RA, and may contribute to our knowledge of how self-tolerance is lost in autoimmune diseases. In addition, determination of transcriptional signatures may lead to the selection of biomarkers for the early detection of RA. Some reports in humans and murine models have identified transcriptional profiles associated with the development of some diseases (10, 11,12). For instance, in patients with arthralgia that latter progressed to RA, Verweij and colleagues [11] identified a transcriptional profile mainly composed of genes associated with Interferon (IFN)-mediated immunity, cytokine/chemokine activity and hematopoiesis. Moreover, another study showed that five interleukins were increased in the serum of subjects that later developed RA [9]. Nevertheless, the molecular events that trigger the development of RA in healthy subjects at high-risk are still not clear.

Our group recently reported that 58% of healthy relatives of RA patients develop the disease during the following 5-year period after the serum levels of anti-citrullinated protein antibodies (anti-CCP) >25 UI/ml were detected. Thus, the detection of anti-CCP antibodies allows the identification of healthy subjects at high risk of developing RA [12]. In this context, the aim...
of this study was to assess the differences in transcriptional profiles between healthy subjects at high-risk of RA and patients with early-RA by means of microarray analysis followed by qPCR validation. Analysis of the differentially expressed pathways is explored and discussed in the context of RA physiopathology and the potential of several genes as biomarkers is also analyzed.

**Material and methods**

**Subjects**

This is a cross-sectional study to assess the differences in transcriptional profile between three groups of subjects described as follows: The group 1 (early-RA; RA) was comprised of RA patients classified according to the criteria of the EULAR/ACR 2010 (26), all of them with less than 1 year of disease evolution (defined from the date of the first swollen joint) and whom were off disease-modifying anti-rheumatic drugs (DMARDs) or glucocorticoids, but were attending a secondary-care rheumatologic clinic as outpatients. Group 2 (high-risk RA; ACCP+) (1) was comprised of healthy first-degree relatives of patients with established RA. All members in this group were older than 18 years, positive to anti-CCP, with no arthritis as per history or any other rheumatic disease. Group 3 (low risk controls; ACCP-) comprised healthy first-degree relatives of patients with established RA. All of them were older than 18 years and anti-CCP negative with no arthritis or any other rheumatic disease as it was described before [13]. All subjects were clinically examined in an independent way by two different certified rheumatologists. In all subjects, anti-CCP antibody levels were assessed using a second-generation IgG anti-CCP2 (ELISA, CCPlus Immunoscan kit, Euro Diagnostica AB, Switzerland) whereas IgM rheumatoid factor (RF) was assessed using nephelometry. According to manufacturer’s instructions, anti-CCP2 levels were considered positive if >25 IU/ml.

The Ethics Committee of the “Instituto Mexicano del Seguro Social” approved the protocol (IMSS, registry number R-20013-785-009). All of the participants enrolled in the study signed a written informed consent letter prior to any procedure.

**Collection of the biological material**

For microarray analysis (discovery set) the following number of subjects was included: 6 ACCP+, 7 early RA and 7 ACCP- group. For the validation set a total of 34 subjects were included (which comprised also the discovery set) as follows: twelve ACCP+, twelve from the reference group (ACCP-), and ten from the early-RA group (RA), all of which were similar in age and gender. Clinical and serologic variables of the studied subjects are shown in Table 1. Individual serum samples were obtained and stored at −20°C until analysis. Whole blood

<table>
<thead>
<tr>
<th>Table 1. Clinical and serologic variables of the studied subjects.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Healthy Relatives</strong></td>
</tr>
<tr>
<td>Anti-CCP2+ N = 12</td>
</tr>
<tr>
<td>Age, yrs., mean ± SD</td>
</tr>
<tr>
<td>Female (%)</td>
</tr>
<tr>
<td>Anti-CCP2 (U/ml, mean ± SD)</td>
</tr>
<tr>
<td>FR (IU/ml, mean ± SD)</td>
</tr>
<tr>
<td>DAS28</td>
</tr>
</tbody>
</table>

*Indicates significant differences for the anti-CCP2- group.

https://doi.org/10.1371/journal.pone.0194205.t001
samples were collected using Vacutainer tubes with EDTA (Becton-Dickinson, USA). Samples were then mixed with 1 ml of R NAlater (Invitrogen, USA), homogenized and frozen at −70˚C until use. Additional samples for Receiver Operating Characteristic (ROC analysis) were included for validation purposes; a subset of genes was evaluated in these samples as potential biomarkers for RA progression.

RNA extraction and cDNA synthesis

Total RNA was extracted using a standardized protocol in our laboratory. Briefly, blood samples were thawed to room temperature and mixed with a volume of TRIzol (Invitrogen, USA). Samples were then mixed with one-tenth of chloroform and centrifuged at 13,000 rpm during 15 min at 4˚C to obtain an aqueous phase with the nucleic acids material. One volume of 70% ethanol was added to the aqueous phase and homogenized. Samples were then processed in QIAmp RNA blood mini-columns (Qiagen, USA) according to the supplier’s instructions [14]. RNA concentration and integrity was evaluated in BioAnalyzer 2100 with the RNA 6000 Nano kit (Agilent Technologies, USA). Only samples with an RNA integrity number (RIN) >8 were considered for the synthesis of complementary Cy3-labeled copy RNA (cRNA). For RT-qPCR determinations 2.5 μg of total RNA from each sample were converted to cDNA using the Superscript II enzyme (Invitrogen, USA), following the supplier’s instructions [15].

Hybridization and analysis of the Agilent GE 4X44 Expression Microarrays

Transcriptional analysis profiles from Cy3-cRNA samples were assessed using seven samples from the early RA patients group, six samples from ACCP+ group, and seven samples from ACCP- group. Microarray hybridization was performed using the Agilent 4X44K platform containing a total of 27,958 genes (Agilent Technologies, USA). Two hundred ng of total RNA were processed with one-color marker protocol, and sample microarray hybridization was carried out following the supplier’s instructions (Agilent Technologies, USA). Mean fluorescence intensities (MFI) from microarrays readings were obtained by use of the Agilent G4900DA SureScan Microarray Scanner (Agilent Technologies, USA). The Agilent Feature Extraction program (Agilent Technologies, USA) was used for individual samples data extraction. A non-supervised analysis of the data was generated with the Agilent GeneSpring software (12.6 ver, Agilent Technologies, USA). Fold change of induced and/or repressed genes between groups was determined by the GeneSpring program through matrix analysis (cluster analysis) and analysis of variance (ANOVA) using a linear threshold of 2 and a P value of less than 0.05 with the Benjamini Hochberg false discovery rate correction. We used Gene Ontology (GO) and Pathway analysis of the Agilent GeneSpring Software platform (Agilent Technologies, USA) to identify the biological function and interaction cascades of identified genes with a significance value for such GO association of P<0.05. Data is available at Gene Expression Omnibus under GSE100191.

Gene validation and amplification by real-time PCR

To corroborate microarray results by qPCR, we considered some of the up-regulated genes in the RA group (early RA), which were associated to innate immune response, humoral innate immune response and inflammatory response. Nine induced genes were selected and primers sequences for qPCR were designed using the Roche online platform [16], (S1 Table). Fifty ng of cDNA per gene in a total volume of 10 μl per-reactions were run by triplicate for each sample. Amplification of cDNA samples was carried out using the SsoFast enzyme EvaGreen (BioRad, USA). We used 45 cycles consisting of 5 seconds of denaturalization at 95˚C, 10 seconds
of alignment at 60˚C, and 10 seconds of extension at 72˚C. The qPCR was conducted employing Roche LightCycler® 480 (Roche Diagnostics, USA). Relative expression of each gene was evaluated by the previously described \(2^{-\Delta\Delta Ct}\) formula [17]. Normalization of qPCR results was done with the housekeeping gene HPRT.

**Statistical analysis**

For statistical analysis in qRT-PCR experiments, GraphPad Prism 5.0 statistical software was used. Significant differences between groups were determined by one-way ANOVA with Tukey post-hoc test if normal distribution was determined, and Kruskal-Wallis test with Dunn post hoc test for non-parametric data. Categorical variables were compared utilizing the \(\chi^2\) test. A ROC curve analysis was performed to evaluate the potential of a subset of genes as biomarkers, AUC (area under the curve) and \(p\) values are reported. In all cases the Two-tailed \(p\) values of <0.05 were considered significant.

**Results**

**Gene expression profiles of ACCP- and relatives ACCP+ compared to early RA patients**

In order to identify the molecular processes associated with RA development and the loss of tolerance, a discovery phase was undertaken in a carefully selected group of individuals by means of microarray data analysis. We identified which genes were differentially expressed between patients with early RA and high risk ACCP+ individual. We identified changes in the transcriptional profiles between the ACCP+ and RA groups versus the group of ACCP- relatives with an unsupervised analysis of the microarray data and a cluster analysis. From the set of 27,958 gene sequences present in the microarray, a specific transcriptome was identified for each group (Fig 1). Disease progression was associated to a list of 551 Up-regulated genes (URG) (S2 Table). Interestingly, >4,000 down-regulated genes (DRG) (S3 Table) were exclusively observed in the early RA group versus the low risk ACCP- group. Considering that subjects with RA develop autoantibodies several years before the disease onset, the differences in transcriptional patterns are important for the understanding of the preclinical autoimmune phase of RA. For this purpose, comparisons between the high-risk ACCP+ group versus the low-risk ACCP- group identified a transcriptome of 876 URG and 7,531 DRG (S4 and S5 Tables). The capacity of such gene expression patterns to categorize individuals into such groups can be visualized in a principal Component Analysis (PCA), which also confirms the transcriptional differences among the three studied groups (Fig 2).

**Differentially expressed genes between ACCP+ and RA**

ACCP antibodies are highly specific for RA and are considered good predictors of the development of RA. To identify the variation in gene expression of the full dataset and to search genes associated with the increased risk to develop RA, a Venn diagram was constructed using the data from the preclinical stage group (ACCP+) and the early RA patients (Fig 3). Shared genes between RA and ACCP+ were identified: 0.2% (17 genes) was URG, mainly related to signal transduction, transcription factors and cytoskeleton, and 35% were DRG genes associated to innate immune response, cell growth, signal transduction and chemical response (full list in S6 and S7 Tables).

Furthermore, 0.4% was URG in the ACCP+ but DRG in RA (S8 Table); A similar 0.4% of the gene repertoire was URG in RA patients but DRG in ACCP+ (S9 Table). Noteworthy, out of the total number of differentially expressed genes, 35% are DRG shared between the high
risk ACCP+ and early RA patients. However, 41% of the DRG were exclusive to the ACCP+ group, and only 9.4% of the DRG were exclusive to the early RA patients. Likewise, 8.4% are URG that belongs specifically to ACCP+ group whereas 5% of the URG are specific to the early RA patients.

Identification of functional categories of the URG in ACCP+ and RA

In order to identify the functional categories of the differentially expressed genes and to have a better understanding of their participation in the physiopathology of RA, a gene ontology (GO) analysis was carried out. As shown in Table 2 a subgroup of up-regulated genes shared between the high risk ACCP+ and early RA patients that are associated to the establishment/development of the disease and the loss of tolerance were identified. These shared gene belong to the innate immune response, IFN type I signature and to leukocyte activation/migration, among others. A full list of genes in each associated biological process in the high-risk group and the early RA patients can be found in S10 and S11 Tables respectively. In particular, GO analysis of the genes overexpressed in RA patients shows that they are associated to inflammation, leukocyte activation, migration and differentiation as well as to hormones and cytokines. Furthermore, the distinctive transcriptome of RA patients is associated to cellular response to
chemical stimuli, stress response and genes associated to regulation of metabolic processes. Also, some of the URG fall under categories such as signal transduction and cell surface receptor signaling pathways (S11 Table).

**Confirmation of microarray data by qPCR**

Given the wide descriptive nature of microarray data, to provide stronger evidence that such massive data are consistent (according to the MIAME guidelines) we used qPCR to confirm microarrays results. For this purpose, additional samples were included as a validation set (including those used in the microarray hybridization experiments). A subset of 9 randomly selected up regulated genes from S11 Table were analyzed by RT-qPCR. Similar results to those obtained with microarray data analysis are shown in Table 3, showing a good overall concordance between the two methods and further confirming the validity of the microarray analysis so far.
Pathway analysis shows a specific set of genes in RA and a shared set of Up-regulated genes in ACCP+ and RA

To associate biological functions of the up regulated genes to define the active pathways that are regulated during the progression process from ACCP+ into RA, we elaborated an overall picture of the biological interactions of the transcriptomes in the context of the studied groups. In Table 4, we show that 8 signaling pathways are specific for early RA URG. Importantly these pathways include genes whose products are involved in the metabolism of phosphatidylinositol, cellular and migration processes as well as immune response against intracellular pathogens such as viruses. The signal pathways deduced from URG in high-risk individuals (ACCP+) and early RA patients were analyzed to identify the molecular interaction between induced genes in our groups. Nineteen induced pathways deduced from URG genes in ACCP+ high-risk individuals were identified (S12 Table). Our analysis shows the presence of induced pathways shared between the ACCP+ group and RA patients (S13 Table). These pathways are mainly related to the WNT signaling pathways, regulation of cytokines synthesis, clotting and the complement system, TCR signaling pathway and the Type I interferon-signaling pathway.

Assessment of genes from the type I interferon pathway, WNT pathway and inflammation as biomarkers for RA

Once that the most relevant pathways were identified, several genes were selected based on their values of fold change (under GSE100191 in NCBI, NIH, Gene Expression Omnibus).
Table 2. Biological processes in ACCP+ and rheumatoid arthritis (RA) groups of up-regulated genes by Gene ontology (GO) analysis.

<table>
<thead>
<tr>
<th>Up-regulated GO in relatives with ACCP+</th>
<th>N, entities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Innate immune response</td>
<td>33</td>
</tr>
<tr>
<td>Interferon type 1 response</td>
<td>5</td>
</tr>
<tr>
<td>Interferon gamma response</td>
<td>10</td>
</tr>
<tr>
<td>Leukocyte activation</td>
<td>17</td>
</tr>
<tr>
<td>Leukocyte migration</td>
<td>12</td>
</tr>
<tr>
<td>Myeloid lineage differentiation</td>
<td>12</td>
</tr>
<tr>
<td>Leukocyte differentiation</td>
<td>10</td>
</tr>
<tr>
<td>Purine-containing compound catabolic process</td>
<td>24</td>
</tr>
<tr>
<td>Cellular response to cytokine stimulus</td>
<td>25</td>
</tr>
<tr>
<td>Cellular response to organonitrogen compounds</td>
<td>18</td>
</tr>
<tr>
<td>Cellular response to hormones</td>
<td>20</td>
</tr>
<tr>
<td>Chemotaxis</td>
<td>23</td>
</tr>
<tr>
<td>Cellular response to insulin stimulus</td>
<td>10</td>
</tr>
<tr>
<td>Cellular response to chemical stimulus</td>
<td>88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Up-regulated GO in patients with RA,</th>
<th>N, entities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inflammatory response</td>
<td>8</td>
</tr>
<tr>
<td>Cell projection organization</td>
<td>21</td>
</tr>
<tr>
<td>Signal transduction</td>
<td>89</td>
</tr>
<tr>
<td>Cellular response to chemical stimulus</td>
<td>41</td>
</tr>
<tr>
<td>Defense response</td>
<td>27</td>
</tr>
<tr>
<td>Immune response</td>
<td>19</td>
</tr>
<tr>
<td>Cellular response to stress</td>
<td>29</td>
</tr>
<tr>
<td>Regulation of metabolic process</td>
<td>122</td>
</tr>
<tr>
<td>Positive regulation of immune system process</td>
<td>17</td>
</tr>
<tr>
<td>Cell surface receptor signaling pathway</td>
<td>49</td>
</tr>
<tr>
<td>Cellular protein modification process</td>
<td>56</td>
</tr>
<tr>
<td>Brain development</td>
<td>21</td>
</tr>
<tr>
<td>Cellular nitrogen compound metabolic process</td>
<td>72</td>
</tr>
</tbody>
</table>

Table 3. Relative expression by RT-qPCR of 9 inflammatory up-regulated genes.

<table>
<thead>
<tr>
<th></th>
<th>First-degree relatives ACCP-(n = 12)</th>
<th>First-degree relatives ACCP+(n = 12)</th>
<th>RA(n = 7)</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCL2</td>
<td>0.45±0.36</td>
<td>0.48±0.43</td>
<td>0.87±0.30</td>
<td>0.0593</td>
</tr>
<tr>
<td>SERPINGB9</td>
<td>0.45±0.40*</td>
<td>0.50±0.20*</td>
<td>1.02±0.53</td>
<td>0.0075</td>
</tr>
<tr>
<td>SERPING1</td>
<td>0.84±0.91</td>
<td>0.96±1.25</td>
<td>0.25±0.32</td>
<td>0.3059</td>
</tr>
<tr>
<td>SNCA</td>
<td>0.37±0.28</td>
<td>0.72±0.74</td>
<td>1.32±0.88</td>
<td>0.1560</td>
</tr>
<tr>
<td>MS4A1</td>
<td>0.31±0.34*</td>
<td>0.40±0.24*</td>
<td>1.45±0.88</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>ETS1</td>
<td>0.53±0.47*</td>
<td>0.61±0.35*</td>
<td>1.3±0.38</td>
<td>&lt;0.011</td>
</tr>
<tr>
<td>EGR1</td>
<td>0.51±0.50*</td>
<td>1.61±2.03*</td>
<td>5.01±4.9</td>
<td>0.0042</td>
</tr>
<tr>
<td>CX3CL1</td>
<td>5.66±7.49</td>
<td>1.79±2.04</td>
<td>0.07±0.09</td>
<td>0.0460</td>
</tr>
<tr>
<td>MEF2A</td>
<td>0.3292±0.31*</td>
<td>0.4733±0.31*</td>
<td>1.048±0.69</td>
<td>0.0041</td>
</tr>
</tbody>
</table>

Values represent mean ± SD; comparison between groups was performed using the Kruskal-Wallis test or ANOVA when necessary. Two-tail p value less than 0.05 was considered significant.

* indicates differences when compared with RA group.
Given that several hurdles still remain in RA diagnosis particularly in the early phases of disease. True diagnosis of RA requires identification of biomarkers related to disease progression. Considering that the up-regulation of genes from the high risk ACCP+ to the early RA patients may reflect the progression of the disease from a pre-clinical stage of the disease, we analyzed the expression of 6 genes to evaluate their utility as biomarkers. First, as shown in Fig 4, significant differences were found in a multiple comparisons test for the next genes: EGR1 and MS4A1 (inflammatory pathway); MXA and IFI6 (Type 1 interferon signature); SOSTDC and WIF1 (Wnt signaling). In order to evaluate the usefulness of some of these genes as biomarkers to identify the RA patients, or high-risk ACCP+ on disease progression, we performed a Receiver Operating Characteristic (ROC) curves. In Fig 5A–5F the ROC analysis showed significant differences compared to the line of no discrimination. Moreover, the AUC for all 6 genes was above 0.8, suggesting their utility as biomarkers. A suggested Cut off point is shown within Fig 5G, with sensitivities similar to those achieved by the immune testing of ACCP and RF (approx. 90%). Remarkably, SOSTDC1 and WIF1, genes of the WNT pathway had the highest AUC (> .92), sensitivity and specificity, and thus may serve as good biomarkers for RA.

Discussion
In this work we evaluated the transcription profiles associated with the development of RA. Thus, using microarray analysis, we compared healthy individuals at high risk (ACCP+), healthy relatives at low risk (ACCP-) and patients with early RA (RA). We identified an important transcriptional arrest on both the ACCP+ and the early RA groups. This suggests that a strong decrease in transcriptional levels precedes the preclinical stage (ACCP+ relatives) and continues up to the development of RA. Similar results had been described previously in a mouse RA progression model, in which mice that develop RA have a greater number of DRG in the previous stages compared to mice that do not develop RA [18]. We hypothesized that the development of RA could be associated with the loss of tolerance promoted by this strong down regulation event, in which many key regulators that control inflammation and suppress pro-inflammatory molecules are reduced. Our data on the increased expression of inflammatory-associated genes support this view. Microarray analysis was confirmed by qPCR. Several major pathways associated with broad regulation of gene expression were identified: Immune response, MAPK signaling, metabolism, Wnt signaling and type I interferon signature. Their possible participation in such phenomenon is discussed below.

MAPK signaling is involved in several cellular communication pathways associated to chronic inflammatory diseases such as RA, Crohn’s disease and psoriasis [19]. In the group of high-risk ACCP+ relatives, we observed an increased expression of genes involved in the
Fig 4. Gene expression of EGR1, MS4A1, MXA, IFI6, WIF1 and SOSTDC1. Gene expression analysis was carried out in cDNA from blood total RNA to assess the relative gene expression profile of selected genes in three groups ACCP-, anti-citrullinated peptide antibodies negative; ACCP+, anti-citrullinated peptide antibodies positive and eRA, early Rheumatoid Arthritis. The graphs depict median ± IQR as descriptive statistics. All relative values used the HPRT expression as reference gene in the $2^{-\Delta\Delta Ct}$ equations. Multiple comparisons tests were made by means of the non-parametric Kruskal-Wallis test. The non-continuous line represents the best cut-off value from ROC analysis. P values of less than 0.05 were considered statistically significant. The * represent significant differences, ** very significant differences and *** extremely significant differences.

https://doi.org/10.1371/journal.pone.0194205.g004
MAPK signal pathway, highlighting PTPN7, DUSP4, STMN1, MKNK1, and FOS. This signal pathway participates directly in the proliferation and differentiation of T and B cells [20, 21], giving rise to a downward trend in the activation of TNFα and IL-1β [20–22]. Other genes, such as PTPN7, which are regulated by ERK1/2 and p38 [23], could be associated with chronic

Fig 5. Receiver operating characteristic (ROC) curves of potential biomarkers. The ROC analysis and curves are shown comparing between ACCP- and early RA subjects for all mRNAs A) ERG1, B) MS4A1, C) MXA, D) IFI6, E) SOSTDC1 F) WIF1, additionally in G) a table summarizing the best cut off points and diagnostic performance of such mRNAs is shown. The line represents the line of no discrimination. P values of less than 0.05 were considered statistically significant.

https://doi.org/10.1371/journal.pone.0194205.g005
proliferation and avoidance of cell death through apoptosis, as has been reported for the PTPN22 gene, whose high expression has been associated directly with the inflammation and with loss of tolerance in RA [24, 25]. Further research is needed to confirm the role of these signaling pathways in the tolerance loss and clinical autoimmunity onset in RA.

As described before, we also observed in the high-risk ACCP+ subjects a set of genes that belongs to the Wnt signals (wingless-type). It has been reported that Wnt signals participate importantly in both synovial inflammation and in bone remodeling [26, 27]. Ligands of the Wnt pathway, such as wnt4, wnt3, and wnt5a, which participate in cellular proliferation, are also highly expressed in RA [28]. wnt5a induce the expression of pro-inflammatory cytokines such as IL-6, IL-8, IL-15, RANKL, fibronectin and MMP3 [29, 30]. Interestingly, it has been demonstrated that the WNT5A gene has a binding site for STAT3, and that its expression is regulated by this transcription factor [31], playing a very important role in the maintenance of a chronic inflammation state, as well as in the destruction of cartilage in arthritis [32–34]. Furthermore, the Wnt/β-catenin signaling pathway regulates the osteoblast and chondroblast metabolism through the DKK1 and Sclerostin, both upstream inhibitors of the Wnt/β-catenin pathway [35]. DKK1 has been proposed as a biomarker of arthritis in early stages of the disease [36]. Our results clearly show a down regulation of the expression of the negative regulators of the WNT signaling pathway, thus suggesting a higher activity in this signaling pathway. In regards to SOSTDC1 (Sclerostin domain-containing protein 1), a member of the family of sclerostin proteins, little information is known about the functions or the regulation of this gene. WIF1 (Wnt inhibitory factor 1) is a member of the WNT regulators and its role in RA pathogenesis is unknown but it has been suggested that might be implicated in cartilage balance and bone anabolism [37]. To our knowledge, this is the first time that WNT regulators are proposed to be used as biomarkers. Further investigation is needed to clarify the role of WNT signaling in the early stages of the disease and for validation of such biomarkers.

The participation of interferon has been widely reported in autoimmune diseases [38] such as Systemic Lupus Erythematous (SLE) and RA [39, 40]. In our group of ACCP+ high levels of expression occur in genes related to both, type I (IFN-α and IFN-β) and type 2 interferon (IFN-γ). This set includes genes that respond to IFN gamma, such as SOCS3, SOCS1, and MT2A. Those genes have been implicated as important negative regulators of the proliferation and release of inflammatory cytokines such as IL-6 [41, 42]. Recently our group found an important association between mRNA gene expression of the type I interferon signature and the production of autoantibodies, providing strong evidence of the participation of these signaling pathways in the physiopathology of RA even before symptoms onset [13]. Particularly, MXA (Interferon-induced dynamin-like GTPase) and IFI6 (Interferon alpha-inducible protein 6) show a modest potential to be used as biomarkers of early RA. Further research is needed in experimental models of RA to establish a causal link between activation of the type I interferon signature and the production of antibodies and B cell activation.

It is widely known that several cytokines such as tumor necrosis factor alpha (TNF-α), Interleukin (IL)-6, IL-17, IL-2 and others play a pivotal role in the RA inflammatory response [41, 43–49]. Our microarray data is consistent with this widely described RA pathological phenomenon. However, little is known regarding the expression of IL-2, IL-7, and IL-9 in preclinical stages of the disease and in promoting inflammation [50, 51]. Although, recent reports suggest that IL-2 production increases in seropositive arthralgia [52], suggesting that it might be associated to the early stage of the inflammatory process in RA. All these molecules are increased in the established disease [50, 51, 53].

MS4A1 (CD20, also known as membrane-spanning 4-domains subfamily A member) has an increased expression in sinovium and correlates with an increased erosion of bone structures in very early RA [54]. This suggests that the elevation of such molecules could be
associated with the expansion of B cell populations responsible for autoantibodies production. This, given that ACPAs have been associated with bone erosions even before clinical symptoms onset [55]. To our knowledge this is the first time that gene expression of MS4A1 is measured in whole blood and proposed as diagnostic biomarker.

EGR-1 is a mammalian transcription factor. Several studies suggest that this molecule is involved in the transcription of metalloproteinases and their specific inhibitors. Also, several reports indicate that EGR1 show an increased expression in sinovial fibroblasts [56] and in infiltrating mononuclear cells in the sinovium regulating proinflammatory molecules production. This depends on the target cells. In mononuclear cells a regulation of PGE2 mediated production of TNF-a [57] was reported. In chondrocytes EGF-1 mediates a decreased expression of extracellular matrix proteins [58] and therefore, it could be regulating the matrix degradation of the sinovium.

The study has several limitations. It is descriptive in nature and the pathways with differential expression need further experimental corroboration. Animal model experiments exploring several candidate molecules and pathways are going to define the role of several of these signaling pathways and their role in RA. The evidence provided in ROC analysis of candidate biomarkers is limited but provides an encouraging prospect for a bigger sample size necessary to verify the utility of the identified genes as biomarker of early stages of RA. In summary, we describe some novel and uncharacterized pathways that could be implicated in the massive down regulation of gene expression that trigger the development of RA. Moreover, from such pathways we identified and evaluated a set of distinctive genes with increased or impaired expression that correlate with the progression of RA with good values of sensitivity and specificity, providing evidence of their potential use as biomarkers of RA. Further studies will be needed to clarify the functional roles of these pathways in the physiopathology of RA as well as to assess the potential use of the identified genes as biomarkers for the early stages of the disease.

Supporting information

S1 Table. Oligonucleotide sequences of the sense and antisense primers for realtime PCR Analysis.

(PDF)

S2 Table. Up regulated genes in patients with RA vs relatives ACCP-.

(PDF)

S3 Table. Down regulated genes in patients with RA vs relatives ACCP-.

(PDF)

S4 Table. Up regulated genes in relatives with ACCP+ vs relatives ACCP-.

(PDF)

S5 Table. Down regulated genes in relatives with ACCP+ vs relatives with ACCP-.

(PDF)

S6 Table. Up regulated genes in AR and ACCP+ groups according Venn diagram.

(PDF)

S7 Table. Down regulated genes in AR and ACCP+ groups according Venn diagram.

(PDF)
S8 Table. Down regulated genes in AR and up regulated genes in ACCP+ groups according Venn diagram.
(PDF)

S9 Table. Up regulated genes in AR and down regulated genes in ACCP+ groups according Venn diagram.
(PDF)

S10 Table. Biological function of the 2 fold change Up-regulated genes according GO analysis in relatives with ACCP+.
(PDF)

S11 Table. Biological function of the 2 fold change Up-regulated genes according GO analysis in RA.
(PDF)

S12 Table. Specific induced pathways in relatives with ACCP+ compared to relatives ACCP-.
(PDF)

S13 Table. Induced pathways in preclinical and clinical AR represented by relatives with ACCP+ and AR patients compared to relatives ACCP-.
(PDF)

Acknowledgments
The authors are grateful to the patients, physicians, laboratory personnel, and healthy volunteers who participated in the study, as well as for the support for the reading of the microarrays provided by Dr. Carlos Córdova-Fletes, Research Professor of the Autonomous University of Nuevo León (UANL), Department of Cytogenomics and Microarrays, and the support of the personnel of the Unit of Arthritis and Rheumatism of Guadalajara City, Jalisco, Mexico.

This project was financed with the support of the “Financial Competition for the Development of Research Protocols and Technological Development on Priority Health Themes, Instituto Mexicano del Seguro Social (IMSS), 2013 Summons”, with number: FIS/IMSS/PROT/PRIIO/13/028. Macías-Segura N. was scholarship recipient of CONACyT (Consejo Nacional de Ciencia y Tecnología), México for the scholarship 321165/ CVU329223.

Author Contributions
Conceptualization: C. Ramos-Remus, J. A. Enciso-Moreno.

Data curation: M. Gomez-Moreno.


Funding acquisition: J. A. Enciso-Moreno.

Investigation: J. D. Castillo-Ortiz.


Project administration: J. A. Enciso-Moreno.

Writing – original draft: N. Macías-Segura, J. E. Castañeda-Delgado.


References


