A comprehensive and centralized database for exploring omics data in Autoimmune Diseases

Jordi Martorell-Marugán [1,2], Raúl López-Domínguez [1], Adrián García-Moreno [1], Daniel Toro-Domínguez [1,3], Juan Antonio Villatoro-García [1], Guillermo Barturen [3], Adoración Martín-Gómez [4], Kevin Troule [5], Gonzalo Gómez-López [5], Fátima Al-Shahrour [5], María Peña-Chilet [6,7,8], Joaquín Dopazo[6,7,8,9], Víctor González-Rumayor [2], Marta E. Alarcón-Riquelme [3,10] and Pedro Carmona-Sáez [1,11*]

[1] Bioinformatics Unit. GENYO. Centre for Genomics and Oncological Research: Pfizer / University of Granada / Andalusian Regional Government, PTS Granada, 18016, Granada, Spain.
[2] Atrys Health S.A., Barcelona, Spain.
[3] Genetics of Complex Diseases. GENYO. Centre for Genomics and Oncological Research: Pfizer / University of Granada / Andalusian Regional Government, PTS Granada, 18016, Granada, Spain.
[4] Nephrology Units. AADEA: Asociación Andaluza de Enfermedades Autoinmunes. Hospital de Poniente, 04700. Almería, Spain.
[5] Bioinformatics Unit. Spanish National Cancer Center, CNIO, Madrid, Spain.
[6] Clinical Bioinformatics Area, Fundación Progreso y Salud (FPS), CDCA, Hospital Virgen del Rocío, 41013, Sevilla, Spain.
[7] Bioinformatics in Rare Diseases (BiER). Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), FPS, Hospital Virgen del Rocio. 41013. Sevilla, Spain.
[8] Computational Systems Medicine, Institute of Biomedicine of Seville (IBIS), Hospital Virgen del Rocio. 41013. Sevilla. Spain
[9] INB-ELIXIR-es, FPS, Hospital Virgen del Rocío, Sevilla, 42013, Spain.
[10] Unit of Chronic Inflammatory Diseases, Institute of Environmental Medicine, Karolinska Institutet, 17177, Stockholm, Sweden.
[11] Department of Statistics. University of Granada, 18071, Granada, Spain.
*Correspondence: pedro.carmona@genyo.es

Summary

Autoimmune diseases are heterogeneous pathologies with difficult diagnosis and few therapeutic options. In the last decade, several omics studies have provided significant insights into the molecular mechanisms of these diseases. Nevertheless, data from different cohorts and pathologies are stored independently in public repositories and a unified resource is imperative to assist researchers in this field. Here, we present ADEx (https://adex.genyo.es), a database that integrates 82 curated transcriptomics and methylation studies covering 5609 samples for some of the most common autoimmune diseases. The database provides, in an easy-to-use environment, advanced data analysis and statistical methods for exploring omics datasets, including meta-analysis, differential expression or pathway analysis.

Introduction

Autoimmune diseases (ADs) are a group of complex and heterogeneous disorders characterized by immune responses to self-antigens leading to tissue damage and dysfunction in several organs. The pathogenesis of ADs is not fully understood, but both environmental and genetic factors have been linked to their development (Salaman, 2003). Although these disorders cause damage to different organs and their clinical outcomes vary between them, they share many risk factors and molecular mechanisms (Jörg et al., 2016). Some examples of ADs are systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), Sjögren's syndrome (SjS), systemic sclerosis (SSc), considered systemic autoimmune diseases (SADs) and type 1 diabetes (T1D), which is considered an organ-specific autoimmune disease. Most of these diseases are classified as rare given their prevalence, but altogether ADs affect up to 3 % of the population considering conservative estimates (Cooper and Stroehla, 2003).

In ADs patients, the pathology is developed during several years but it is only detected when tissue damage is significant. For that reason, early diagnosis is important and complicated. Additionally, some ADs often show a non-linear outcome that alternates between active and remission stages thus making their study even more difficult. Despite huge efforts have been made to develop ADs biomarkers and therapies, these do not fit for every patient and their clinical responses differ greatly (Barturen et al., 2018).

During the past decade, the use of omics technologies has provided new insights into the molecular mechanisms associated with the development of ADs, opening new scenarios for biomarkers and treatments discovery (Kim et al., 2014). In this context, it is remarkable the characterization of the type I interferon (IFN) gene expression signature as a key factor in the pathology of some SADs, especially in SLE and SjS (Thorlacius et al., 2018), which has improved our knowledge of the underlying molecular mechanisms and has opened new therapeutic strategies based on blocking the pathways related to this signature.

Regardless of the large amount of omics studies describing new biomarkers and therapeutic strategies in ADs (Arriens and Mohan, 2013; Ferreira et al., 2014; Teruel et al., 2017; Xie et al., 2018), in most

cases these biomarkers are not consistent across different studies or have not fully accomplished their diagnostic goals. Indeed, the widely studied IFN signature is highly variable between patients (Rönnblom and Eloranta, 2013) and it is associated to differences in response to treatments which target it, as has been reported for example in the phase-II results of Sifalimumab clinical trial for SLE patients (Khamashta et al., 2016). In addition, in most of the cases, biomarkers are defined from the analysis of a single type of omic data (commonly gene expression), but multi-omics data integration can provide a more complete understanding of molecular mechanisms and more robust and biologically relevant biomarkers.

Most of the omics datasets generated from different cohorts and studies in ADs published to date have been deposited and are available in public repositories such as Gene Expression Omnibus (GEO) (Edgar et al., 2002) or ArrayExpress (Kolesnikov et al., 2015). Although all these valuable data can be used in retrospective analyses in order to generate new knowledge and accelerate drug discovery and diagnosis, it is not easy to compare neither to integrate available data because they are generated from different platforms and/or processed with different analytic pipelines. In this context, there are great efforts from the bioinformatics community to develop standardized data analysis workflows and resources that facilitate data integration and reproducible analysis. For example, Lachmann et al. (Lachmann et al., 2018) have recently reprocessed a large collection of raw human and mouse RNA-Seq data from GEO and Sequence Read Archive (SRA) using a unified pipeline and they have developed the ARCHS4 as a resource to provide direct access to these data through a web-based user interface. Other singular projects such as The Cancer Genome Atlas (TCGA) (Weinstein et al., 2013) or the Genotype-Tissue Expression project (GTEx) (Lonsdale et al., 2013) provide also large and homogeneously processed datasets for tumor samples and human tissues respectively. These unprecedented resources motivate the development of applications and data portals to help researchers gather information with the aim of improving diagnosis and treatment in multiple diseases, most notably in cancer research, where such information is actually being used in the clinical practice (Jang et al., 2018).

Despite such enormous potential, in the context of ADs there is a lack of a centralized and dedicated resource that facilitates the exploration, comparison and integration of available omics datasets. This is

indeed an area in which this type of application would be tremendously beneficial, given that the low prevalence of each individual disease makes difficult the recruitment of large patients cohorts (Barturen et al., 2018).

To bridge this gap, in this work we have compiled and curated most of the publicly available gene expression and methylation datasets for five ADs: SLE, RA, SjS, SSc and T1D. To this end, we have developed and applied homogeneous pipelines from raw data and we developed ADEx (Autoimmune Disease Explorer), a data portal where these processed data can be downloaded and exploited through multiple exploratory and statistical analyses. ADEx facilitates data integration and analysis to potentially improve diagnosis and treatment of ADs.

Results

Data collection and processing

ADEx contains data from 5609 samples. We have processed 82 expression and methylation datasets from case-control studies for SLE, RA, SjS, SSc and T1D diseases (see Table 1 for a summary and Supplementary Table 1 for complete information about all included datasets). We have manually curated all metadata in order to standardize the nomenclature of phenotypes, cell types, etc. from different studies and discard samples or datasets that do not meet the selection criteria (see Methods

Table 1. Summary of accessible studies and samples by disease and data type in ADEx.

	Expression	**Methylation**	**Total**
Disease	**Datasets - Samples**	**Datasets - Samples**	**Datasets - Samples**
SLE	20 - 2053	13 - 628	33 - 2626
RA	17 - 1122	3 - 835	20 - 1957
SjS	9 - 400	1 - 29	10 - 429
SSc	5 - 229	1 - 37	6 - 229
T1D	11 - 176	2 - 100	13 - 276

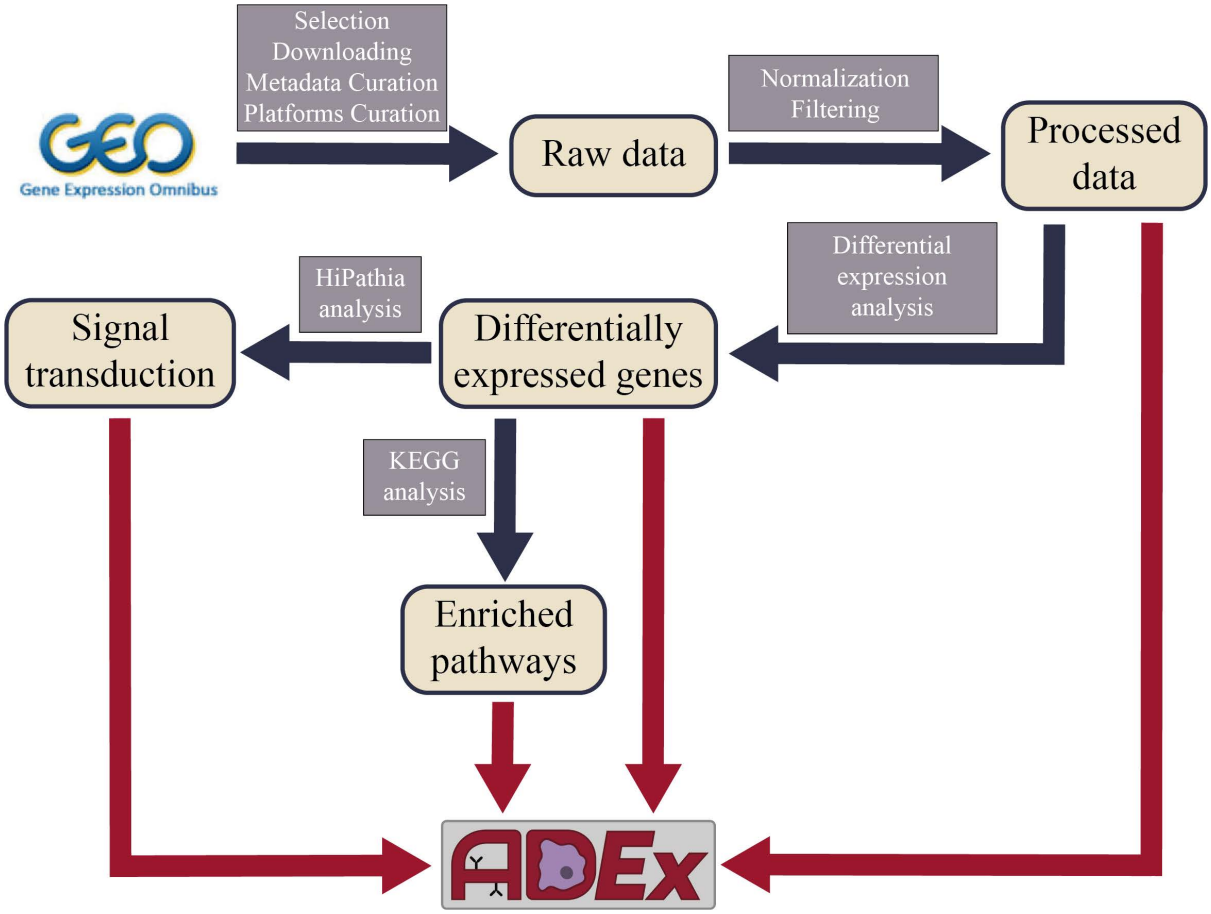

Figure 1. Processing pipeline for ADEx data. Black arrows indicate intermediate processing steps. Red arrows indicate the inputs to ADEx application.

section). In addition, we have prepared five different pipelines to process data for each platform (RNA-Seq, Affymetrix and Illumina gene expression microarrays, and Illumina methylation arrays 27K and 450K).

All these workflows are written in R language and are publicly available in GENyO Bioinformatics Unit GitHub (https://github.com/GENyO-BioInformatics/public/tree/master/ADEx). The processed datasets are available from the Download Data section in the application. Figure 1 contains an overview of the different steps performed to prepare the data for ADEx application.

The ADEx application

ADEx data portal can be used to download and analyze the processed data. ADEx is freely available at https://adex.genyo.es. The tool is divided in 6 different sections arranged in different tabs (Figure 2a).

Section 1: Data overview

Information about the available datasets can be found in both table or pie plot formats in this section. In tables, information about the sample phenotype and their data origin is provided. In pie plots quantitative information is provided regarding the clinical and phenotype information. All this information has been extracted from GEO or from the associated published articles whenever supplied. This information can be presented individually for each dataset or grouped by disease. While a single dataset is being explored, the experiment summary is shown. Users can use this section to identify datasets of their interest to be analyzed in the following sections.

Section 2: Gene Query

This section was created in order to explore the expression and methylation of a specific gene, or the correlation between them, within a single dataset. Users can explore the different gene expression values for each dataset comparing case and control samples with a boxplot. Meanwhile, methylation data is presented at CpG level, so that users can select a region of the gene (e.g. promoter) and the mean methylation value for cases and controls is plotted for every CpG probe contained in the selected region. It has been demonstrated the strong relationship of gene expression and methylation levels (Suzuki and Bird, 2008). That is why, in this section, users can also integrate both expression and methylation values to search for direct or inverse correlations. Finally, gene expression correlation analysis can be performed in order to get insight into the relationship between different genes and to find groups of coexpressed genes.

Section 3: Gene Set Query

Here users can select several datasets and genes in order to explore the Fold-Change (FC) between patients and controls across studies. All datasets from a disease can be automatically selected by clicking the right buttons, or individual studies can be selected by clicking directly on the table. Users can introduce a list of genes to explore their expression, although there are several preloaded gene lists covering the coexpression modules reported by Chaussabel et al. (Chaussabel et al., 2008). These modules consist of sets of coexpressed genes among hundreds of samples from different diseases. Each

transcriptional module is associated to different pathways and cell types, most of them related to the immune system (Chaussabel et al., 2008). See Figures 2b and 2c for an example of this type of analysis.

Section 4: Analyze Dataset

In this section, we focus the analysis on whole datasets instead of individual genes. By default, a heatmap with the expression of the top 50 differentially expressed genes (DEGs) sorted by FDR is displayed. It is also possible to sort them by FC and FDR cutoffs can be applied to both statistics. Additionally, differential expression analysis results can be downloaded as an excel table.

Furthermore, users can also study the KEGG (Kanehisa and Goto, 2000) enriched pathways associated to the dataset selected. These results are precomputed using all the DEGs that have an FDR value below 0.05. A table gathers the significantly enriched KEGG pathways along with their associated hypergeometric test statistics and an interactive plot shows detailed information of the participant genes in the pathway colored according to their FC.

Beyond conventional pathway enrichment methods, we have implemented more sophisticated mechanistic models of cell signaling activity which have demonstrated to be very sensitive in deciphering disease mechanisms (Cubuk et al., 2018; Hidalgo et al., 2017) as well as the mechanisms of action of drugs (Amadoz et al., 2015; Esteban-Medina et al., 2019).

To offer this functionality we have applied HiPathia software (Hidalgo et al., 2017) to gene expression data. This method estimates changes in the activity of signaling circuits defined into different pathways. With this approach, it becomes possible to study in detail the specific signaling circuits altered in ADs within the different signaling pathways. We precomputed this analysis for each dataset and the results are available as tables and interactive reports.

Section 5: Meta-Analysis

ADEx also implements meta-analysis functionalities based on gene expression data to integrate and jointly analyze different and heterogeneous datasets. We implemented a meta-analysis approach to search for biomarkers and common gene signatures across different datasets from the same or different pathologies (Toro-Domínguez et al., 2014a) based on the FCs of each dataset and gene. Datasets have to be selected similarly to Section 3 to launch the meta-analysis.

Section 6: Download data

In this section, users can select one or several datasets and download them. Curated data is obtained with the aim of performing additional analyses externally to ADEx application.

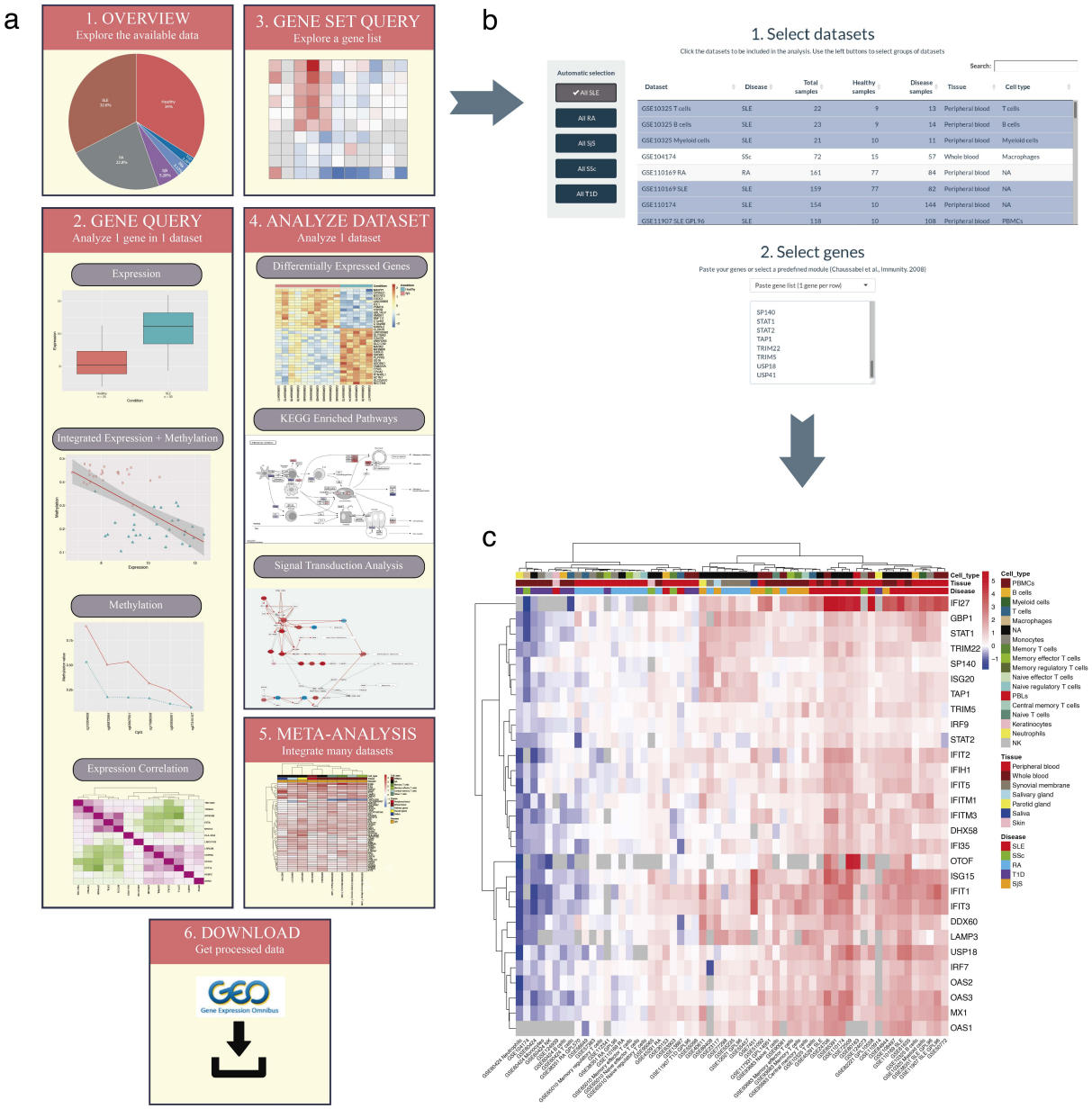

Figure 2. Overview of ADEx application and analysis of IFN signature across diseases. a) ADEx has six main sections. Section 1 provides information about available datasets. In section 2, users can explore expression and methylation for individual genes. Section 3 implements a module to explore data for a gene list, such as gene module or genes from a biological pathway, across several datasets. Section 4 allows researchers to perform analysis on individual datasets retrieving differential expression signatures and pathways and cell signaling enrichment analyses. Section 5 implements meta-analysis methods to integrate multiple datasets in order to define common biomarkers. Section 6 is for data download. **b) Gene Set Query section screenshot.** Datasets and gene set input is shown. Users select data there to plot a heatmap. **c) IFN signature expression generally separates SLE and SjS from other ADs.** Heatmap with the IFN genes generated in ADEx. Color represents the $\log_2$ FC of disease versus healthy samples (red for overexpression and blue for underexpression).

Methods

Data collection

Collection of the datasets included in ADEx was carried out by searching in GEO web page with ADs names as key terms. We filtered the results by study type (expression profiling by array, expression profiling by high throughput sequencing and methylation profiling by array), organism (Homo sapiens) and platform manufacturer (Affymetrix or Illumina).

We downloaded the metadata for these initial datasets with GEOquery (Davis and Meltzer, 2007) R package in order to apply our inclusion criteria and exclude those studies and samples that do not meet them. We only included case-control studies from samples, which were not treated with drugs *in vitro*. Exclusively datasets with available raw data were considered. Studies whose controls and cases belong to different tissues were discarded. We only selected datasets with 10 samples at least. We divided the datasets containing samples from different diseases, platforms, tissues or cell types in subgroups so that these are constant and avoid possible batch effects.

82 datasets containing 5609 samples passed our filtering criteria (see Table 1 for a summary and Supplementary Table 1 for complete information about all included datasets). Then we downloaded their raw data with GEOquery (Davis and Meltzer, 2007). For expression microarrays, we downloaded CEL files and raw text files for Affymetrix and Illumina platforms respectively. For RNA-Seq, we downloaded the *fastq* files from the European Nucleotide Archive. For methylation microarrays, we downloaded raw methylation tables if they were available and *idat* files otherwise.

Metadata curation

GEO does not require submitters to use either a fixed structure or standard vocabulary to describe the samples of an experiment. For that reason, it was necessary to manually homogenize the information provided within all the selected datasets using standardized terms. There are some methods for automatic curation of GEO metadata, but manual curation is still necessary to get high-quality metadata (Wang et al., 2019). This metadata curation was an essential step for the following analyses and permits an easy datasets information exploration.

Platforms curation

We have used a total of 12 different gene expression platforms from microarray and RNA-Seq technologies. Microarray platforms quantify expression levels in probes. In order to match probe identifiers to gene names, platforms annotation files are available from GEO. However, we found that some of these annotation files match probes to inappropriate gene names. On the one hand, some platforms save gene names with errors due to the conversion of gene names such as *MARCH1* or *SEPT1* into dates, a common error that has been reported previously (Ziemann et al., 2016). In these cases, we fixed manually these genes in the annotation files. On the other hand, different platforms use obsolete or simply different synonyms or aliases to refer to the same genes. We used human genes' information from NCBI repository in order to match aliases with actual official gene symbols and substituted them in the platform annotations.

Data processing

Raw data from Illumina expression microarrays were loaded by reading the plain text files. In order to remove background noise, we kept only the probes that had a Detection P-value lower than 0.05 in 10 % of the samples. Then we performed a background correction and quantile normalization (Shi et al., 2010) using *neqc* function from limma package (Ritchie et al., 2015).

CEL files from Affymetrix expression microarrays platforms were loaded to R environment with *affy* package (Gautier et al., 2004). To filter low intensity probes, we removed all probes with an intensity lower than 100 in at least 10 % of the samples. Normalization was carried out computing Robust Multichip Average (RMA) normalization (Irizarry et al., 2003) with *affy* package (Gautier et al., 2004).

For RNA-Seq datasets, *fastq* files were aligned to human transcriptome reference hg38 using STAR 2.4 (Dobin et al., 2013) and raw counts were obtained with RSEM v1.2.31 (Li and Dewey, 2011) with default parameters. Raw counts were filtered using NOISeq R package (Tarazona et al., 2015), removing those features that have an average expression per condition lower than 0,5 counts per million (CPM) and a coefficient of variation (CV) higher than 100 in all conditions. Counts normalization was carried out with TMM method (Robinson and Oshlack, 2010).

We translated microarrays probes identifiers to gene symbols using our curated annotation tables. For those genes targeted by two or more microarray probes, we calculated the median expression values of all their targeting probes. For RNA-Seq, we translated ENSEMBL identifiers to gene symbols using biomaRt package (Durinck et al., 2005, 2009).

Methylation raw data are available in GEO as idat or text files depending on the dataset. Idat files were read with *minfi* package (Aryee et al., 2014), while text files were read in the R environment. In both cases, poorly performing probes with a detection P-value above 0.05 in more than 10 % of samples were removed. Probes adjacent to SNPs, located in sexual chromosomes or reported to be cross-reactive (Chen et al., 2013) were also removed. We normalized the methylation signals using quantile normalization with *lumi* package (Du et al., 2008). Finally, for datasets generated with 450k platform, we applied BMIQ normalization (Teschendorff et al., 2013) using wateRmelon package (Pidsley et al., 2013) in order to correct for the two types of probes contained in this platform.

Differential expression analysis

We performed a differential expression analysis in all datasets independently towards the identification of differential patterns among disease samples and healthy controls. These analyses were performed in different ways depending on the source of data. Gene expression profiles from microarray platforms were carried out by the standard pipeline of limma package (Ritchie et al., 2015). We used lmFit function to fit a linear model to the gene expression values followed by the execution of a t-test by the empirical Bayes method for differential activity (eBayes function). On the other hand, gene expression profiles from RNA-Seq platforms were analyzed by the standard pipeline of DESeq2 package (Love et al., 2014). In both cases, differential expression analysis provided P-values, adjusted P-values by FDR and $\log_2$ FC.

Pathway analysis

Pathway enrichment analysis was precomputed for each expression dataset using differential expression analysis results. We considered DEGs those genes with a FDR lower than 0.05 and we performed hypergeometric tests to check if each pathway contains more DEGs as expected by chance. We used

KEGGprofile 1.24.0 R package to perform this analysis but beforehand we manually updated its dependency, KEGG.db, the database used to perform the statistical test. The pathways were plotted using KEGG mapper tool, Search&Color Pathway, with the genes colored by their FC between case and control samples.

Signaling network analysis

We integrated signaling network analysis applying HiPathia software (Hidalgo et al., 2017) to gene expression data so that changes in the activity of the network from different pathways can be detected. We precomputed this analysis for each gene expression dataset. Firstly, we translated the gene expression matrix and scaled it. Then, we calculated the transduction signal and compared among conditions, cases and controls. Finally, the results were stored in interactive html reports.

Database architecture

Pursuing an optimal data organization and a quick access to all the data in ADEx, we have enabled an internal database with PostgreSQL. We chose this technology since it is open source and it is best suited to the huge dimensionality of omics datasets.

Webtool

ADEx user interface was designed with RStudio Shiny package. The application uses a set of external packages to perform analysis and graphics on demand. Most of the plots are generated with ggplot2 (Wickham, 2009). All the computations in Meta-Analysis section are performed whenever users request them. Biomarkers analysis is performed with Rank Products algorithm integrated in RankProd R package (Del Carratore et al., 2017). The tool runs in our own server with CentOS 7.0 operating system, 16 processors and 32 Gb of RAM memory.

Discussion

Despite that the heterogeneity of ADs is evident, there are common molecular mechanisms involved in the activation of immune responses. In this context, integrative analyses of multiple studies are crucial

to discover shared and differential molecular signatures (Toro-Domínguez et al., 2014b). Nowadays there are many ADs datasets publicly available, but a strong computational knowledge is necessary in order to analyze them properly. With the aim of filling this gap between experimental research and computational biology, interactive easy-to-use software are valuable tools to perform exploratory and statistical analysis without strong computational expertise. This type of tool has been developed for other diseases and has helped to reuse public data and generate new knowledge and hypotheses (Cerami et al., 2012; Díez-Villanueva et al., 2015; Toro-Domínguez et al., 2019).

A resource of this type is urged in the field of ADs to: **1)** Compile available ADs' public data in a single data portal, **2)** Access to integrable data processed with uniform pipelines, and **3)** Perform both individual and integrated analysis interactively. We developed ADEx database to accomplish all those objectives.

As far as we know, ADEx is the first ADs omics database and we expect it to be a reference in this area. During the coming years, ADEx will be expanded including data from more ADs and other omics.

Acknowledgements

We would like to thank all the authors of the datasets included in ADEx. We also would like to thank Alberto Ramírez for his technical support during the implementation of ADEx in our server. This work is part of the JMM's PhD thesis. JMM is enrolled in the PhD program in Biomedicine at the University of Granada, Spain.

JMM is partially funded by Ministerio de Economía, Industria y Competitividad. This work is partially funded by Consejería de Salud, Junta de Andalucía (Grant PI-0173-2017).

References

Amadoz, A., Sebastian-Leon, P., Vidal, E., Salavert, F., and Dopazo, J. (2015). Using activation status of signaling pathways as mechanism-based biomarkers to predict drug sensitivity. Sci. Rep. *5*, 18494.

Arriens, C., and Mohan, C. (2013). Systemic lupus erythematosus diagnostics in the 'omics' era. Int. J. Clin. Rheumatol. *8*, 671–687.

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of

Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369.

Banchereau, R., Hong, S., Cantarel, B., Baldwin, N., Baisch, J., Edens, M., Cepika, A.-M., Acs, P., Turner, J., Anguiano, E., et al. (2016). Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. Cell *165*, 551–565.

Barturen, G., Beretta, L., Cervera, R., Van Vollenhoven, R., and Alarcón-Riquelme, M.E. (2018). Moving towards a molecular taxonomy of autoimmune rheumatic diseases. Nat. Rev. Rheumatol. *14*, 75–93.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404.

Chaussabel, D., Quinn, C., Shen, J., Patel, P., Glaser, C., Baldwin, N., Stichweh, D., Blankenship, D., Li, L., Munagala, I., et al. (2008). A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity *29*, 150–164.

Chen, Y., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J., and Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics *8*, 203–209.

Cooper, G.S., and Stroehla, B.C. (2003). The epidemiology of autoimmune diseases. Autoimmun. Rev. *2*, 119–125.

Crow, M.K. (2014). Type I Interferon in the Pathogenesis of Lupus. J. Immunol. Baltim. Md 1950 *192*, 5459–5468.

Cubuk, C., Hidalgo, M.R., Amadoz, A., Pujana, M.A., Mateo, F., Herranz, C., Carbonell-Caballero, J., and Dopazo, J. (2018). Gene Expression Integration into Pathway Modules Reveals a Pan-Cancer Metabolic Landscape. Cancer Res. *78*, 6059–6072.

Davis, S., and Meltzer, P.S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinforma. Oxf. Engl. *23*, 1846–1847.

Del Carratore, F., Jankevics, A., Eisinga, R., Heskes, T., Hong, F., and Breitling, R. (2017). RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets. Bioinforma. Oxf. Engl. *33*, 2774–2775.

Díez-Villanueva, A., Mallona, I., and Peinado, M.A. (2015). Wanderer, an interactive viewer to explore DNA methylation and gene expression data in human cancer. Epigenetics Chromatin *8*.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinforma. Oxf. Engl. *29*, 15–21.

Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. Bioinforma. Oxf. Engl. *24*, 1547–1548.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinforma. Oxf. Engl. *21*, 3439–3440.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210.

Esteban-Medina, M., Peña-Chilet, M., Loucera, C., and Dopazo, J. (2019). Exploring the druggable space around the Fanconi anemia pathway using machine learning and mechanistic models. BMC Bioinformatics *20*, 370.

Ferreira, R.C., Guo, H., Coulson, R.M.R., Smyth, D.J., Pekalski, M.L., Burren, O.S., Cutler, A.J., Doecke, J.D., Flint, S., McKinney, E.F., et al. (2014). A type I interferon transcriptional signature precedes autoimmunity in children genetically at risk for type 1 diabetes. Diabetes *63*, 2538–2550.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. Bioinforma. Oxf. Engl. *20*, 307–315.

Guo, Q., Wang, Y., Xu, D., Nossent, J., Pavlos, N.J., and Xu, J. (2018). Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies. Bone Res. *6*, 15.

Hidalgo, M.R., Cubuk, C., Amadoz, A., Salavert, F., Carbonell-Caballero, J., and Dopazo, J. (2017). High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. Oncotarget *8*, 5160–5178.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostat. Oxf. Engl. *4*, 249–264.

Jang, Y., Choi, T., Kim, J., Park, J., Seo, J., Kim, S., Kwon, Y., Lee, S., and Lee, S. (2018). An integrated clinical and genomic information system for cancer precision medicine. BMC Med. Genomics *11*.

Jörg, S., Grohme, D.A., Erzler, M., Binsfeld, M., Haghikia, A., Müller, D.N., Linker, R.A., and Kleinewietfeld, M. (2016). Environmental factors in autoimmune diseases and their role in multiple sclerosis. Cell. Mol. Life Sci. *73*, 4611–4622.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

Khamashta, M., Merrill, J.T., Werth, V.P., Furie, R., Kalunian, K., Illei, G.G., Drappa, J., Wang, L., Greth, W., and CD1067 study investigators (2016). Sifalimumab, an anti-interferon-α monoclonal antibody, in moderate to severe systemic lupus erythematosus: a randomised, double-blind, placebo-controlled study. Ann. Rheum. Dis. *75*, 1909–1916.

Kim, H.-Y., Kim, H.-R., and Lee, S.-H. (2014). Advances in systems biology approaches for autoimmune diseases. Immune Netw. *14*, 73–80.

Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T., et al. (2015). ArrayExpress update--simplifying data submissions. Nucleic Acids Res. *43*, D1113-1116.

Lachmann, A., Torre, D., Keenan, A.B., Jagodnik, K.M., Lee, H.J., Wang, L., Silverstein, M.C., and Ma'ayan, A. (2018). Massive mining of publicly available RNA-seq data from human and mouse. Nat. Commun. *9*, 1366.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F.,

Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Nguyen, C.Q., and Peck, A.B. (2013). The Interferon-Signature of Sjögren's Syndrome: How Unique Biomarkers Can Identify Underlying Inflammatory and Immunopathological Mechanisms of Specific Diseases. Front. Immunol. *4*, 142.

Pidsley, R., Y Wong, C.C., Volta, M., Lunnon, K., Mill, J., and Schalkwyk, L.C. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics *14*, 293.

Pollard, K.M., Cauvi, D.M., Toomey, C.B., Morris, K.V., and Kono, D.H. (2013). Interferon-γ and Systemic Autoimmunity. Discov. Med. *16*, 123–131.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Rönnblom, L., and Eloranta, M.-L. (2013). The interferon signature in autoimmune diseases. Curr. Opin. Rheumatol. *25*, 248–253.

Rusinova, I., Forster, S., Yu, S., Kannan, A., Masse, M., Cumming, H., Chapman, R., and Hertzog, P.J. (2013). INTERFEROME v2.0: an updated database of annotated interferon-regulated genes. Nucleic Acids Res. *41*, D1040–D1046.

Salaman, M.R. (2003). A two-step hypothesis for the appearance of autoimmune disease. Autoimmunity *36*, 57–61.

Shi, W., Oshlack, A., and Smyth, G.K. (2010). Optimizing the noise versus bias trade-off for Illumina whole genome expression BeadChips. Nucleic Acids Res. *38*, e204.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat. Rev. Genet. *9*, 465–476.

Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A.D., Nueda, M.J., Ferrer, A., and Conesa, A. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res. *43*, e140.

Teruel, M., Chamberlain, C., and Alarcón-Riquelme, M.E. (2017). Omics studies: their use in diagnosis and reclassification of SLE and other systemic autoimmune diseases. Rheumatol. Oxf. Engl. *56*, i78–i87.

Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., and Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinforma. Oxf. Engl. *29*, 189–196.

Thorlacius, G.E., Wahren-Herlenius, M., and Rönnblom, L. (2018). An update on the role of type I interferons in systemic lupus erythematosus and Sjögren's syndrome. Curr. Opin. Rheumatol. *30*, 471–481.

Toro-Domínguez, D., Carmona-Sáez, P., and Alarcón-Riquelme, M.E. (2014a). Shared signatures

between rheumatoid arthritis, systemic lupus erythematosus and Sjögren's syndrome uncovered through gene expression meta-analysis. Arthritis Res. Ther. *16*, 489.

Toro-Domínguez, D., Carmona-Sáez, P., and Alarcón-Riquelme, M.E. (2014b). Shared signatures between rheumatoid arthritis, systemic lupus erythematosus and Sjögren's syndrome uncovered through gene expression meta-analysis. Arthritis Res. Ther. *16*.

Toro-Domínguez, D., Martorell-Marugán, J., López-Domínguez, R., García-Moreno, A., González-Rumayor, V., Alarcón-Riquelme, M.E., and Carmona-Sáez, P. (2019). ImaGEO: integrative gene expression meta-analysis from GEO database. Bioinforma. Oxf. Engl. *35*, 880–882.

Wang, Z., Lachmann, A., and Ma'ayan, A. (2019). Mining data and metadata from the gene expression omnibus. Biophys. Rev. *11*, 103–110.

Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer Analysis Project. Nat. Genet. *45*, 1113–1120.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag).

Xie, X., Li, F., Li, S., Tian, J., Chen, J.-W., Du, J.-F., Mao, N., and Chen, J. (2018). Application of omics in predicting anti-TNF efficacy in rheumatoid arthritis. Clin. Rheumatol. *37*, 13–23.

Ziemann, M., Eren, Y., and El-Osta, A. (2016). Gene name errors are widespread in the scientific literature. Genome Biol. *17*, 177.

Supplementary Information

Supplementary Table 1. Description of the datasets included in ADEx database. This table contains information about each study included in ADEx, with disease, platform, sample size and reference (if available).

Dataset	Studied disease	Experimental strategy	Platform	Sample size	Reference
GSE10325	SLE	Expression profiling by array	[HG-U133A] Affymetrix Human Genome U133A Array	67	(Hutcheson et al., 2008)
GSE104174	SSc	Expression profiling by high throughput sequencing	Illumina HiSeq 2500 (Homo sapiens)	72	(Moreno-Moral et al., 2018)
GSE108497	SLE	Expression profiling by array	Illumina HumanHT-12 V4.0 expression beadchip	512	NA
GSE110007	SjS	Methylation profiling by array	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	31	(Cole et al., 2016)
GSE110169	SLE, RA	Expression profiling by array	[HG-U219] Affymetrix Human Genome U219 Array	234	NA
GSE110174	SLE	Expression profiling by array	[HT_HG-U133_Plus_PM] Affymetrix HT HG-U133+ PM Array Plate	154	NA
GSE110607	SLE	Methylation profiling by genome tiling array	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	104	(Ulff-Møller et al., 2018)

Dataset	Studied disease	Experimental strategy	Platform	Sample size	Reference
GSE110914	T1D	Expression profiling by high throughput sequencing	Illumina HiSeq 2500 (Homo sapiens)	42	(Vecchio et al., 2018)
GSE112341	T1D	Expression profiling by high throughput sequencing	Illumina HiSeq 2500 (Homo sapiens)	22	(Gao et al., 2019)
GSE117931	SSc	Expression profiling by array, Methylation profiling by genome tiling array	llumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip, Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	74	NA
GSE11907	SLE	Expression profiling by array	[HG-U133A] Affymetrix Human Genome U133A Array [HG-U133B] Affymetrix Human Genome U133B Array	546	(Chaussabel et al., 2008)
GSE12021	RA	Expression profiling by array	[HG-U133A] Affymetrix Human Genome U133A Array [HG-U133B] Affymetrix Human Genome U133B Array	57	(Huber et al., 2008)
GSE124073	SSc	Expression profiling by high throughput sequencing	Illumina HiSeq 2000 (Homo sapiens)	73	(Mariotti et al., 2019)
GSE124939	SLE	Expression profiling by high throughput sequencing	Illumina HiSeq 4000 (Homo sapiens)	72	(Tsoi et al., 2019)
GSE13887	SLE	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	27	(Fernandez et al., 2009)
GSE23117	SjS	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	15	(Greenwell-Wild et al., 2011)
GSE24706	SLE	Expression profiling by array	Illumina HumanWG-6 v3.0 expression beadchip	48	(Li et al., 2011)
GSE27895	SLE	Methylation profiling by array	Illumina HumanMethylation27 BeadChip (HumanMethylation27_270596_v.1.2)	23	(Jeffries et al., 2011)
GSE30153	SLE	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	26	(Garaud et al., 2011)
GSE38351	SLE,RA	Expression profiling by array	[HG-U133A] Affymetrix Human Genome U133A Array [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	74	(Smiljanovic et al., 2012)
GSE40611	SjS	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	49	(Horvath et al., 2012)
GSE42861	RA	Methylation profiling by array	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	689	(Liu et al., 2013)

Dataset	Studied disease	Experimental strategy	Platform	Sample size	Reference
GSE45291	SLE,RA	Expression profiling by array	[HT_HG-U133_Plus_PM] Affymetrix HT HG-U133+ PM Array Plate	805	(Bienkowska et al., 2014)
GSE50772	SLE	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	81	(Kennedy et al., 2015)
GSE51092	SjS	Expression profiling by array	Illumina HumanWG-6 v3.0 expression beadchip	222	(Lessard et al., 2013)
GSE55098	T1D	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	22	(Yang et al., 2015)
GSE55235	RA	Expression profiling by array	[HG-U133A] Affymetrix Human Genome U133A Array	30	(Woetzel et al., 2014)
GSE55457	RA	Expression profiling by array	[HG-U133A] Affymetrix Human Genome U133A Array	33	(Woetzel et al., 2014)
GSE56606	T1D	Methylation profiling by array	Illumina HumanMethylation27 BeadChip (HumanMethylation27_270596_v.1.2)	100	(Rakyan et al., 2011)
GSE56649	RA	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	22	(Ye et al., 2015)
GSE57383	RA	Expression profiling by array	[HT_HG-U133_Plus_PM] Affymetrix HT HG-U133+ PM Array Plate	112	(Rosenberg et al., 2014)
GSE57869	SLE	Methylation profiling by array	Illumina HumanMethylation27 BeadChip (HumanMethylation27_270596_v.1.2)	12	(Hong et al., 2017)
GSE59250	SLE	Methylation profiling by array	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	434	(Absher et al., 2013)
GSE60424	T1D	Expression profiling by high throughput sequencing	Illumina HiScanSQ (Homo sapiens)	134	(Linsley et al., 2014)
GSE61635	SLE	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	129	NA
GSE63903	SSc	Expression profiling by array	Illumina HumanHT-12 V4.0 expression beadchip	14	(Ayano et al., 2015)
GSE65010	RA	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	48	(Walter et al., 2016)
GSE65391	SLE	Expression profiling by array	Illumina HumanHT-12 V4.0 expression beadchip	996	(Banchereau et al., 2016)
GSE71841	RA	Methylation profiling by array	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	24	NA
GSE72509	SLE	Expression profiling by high throughput sequencing	Illumina HiSeq 2500 (Homo sapiens)	117	(Hung et al., 2015)

Dataset	Studied disease	Experimental strategy	Platform	Sample size	Reference
GSE7451	SjS	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	20	(Hu et al., 2007)
GSE77298	RA	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	23	(Broeren et al., 2016)
GSE80183	SLE	Expression profiling by high throughput sequencing	Illumina HiSeq 2000 (Homo sapiens)	16	(Rai et al., 2016)
GSE82221	SLE	Expression profiling by array, Methylation profiling by genome tiling array	Illumina HumanHT-12 V4.0 expression beadchip, Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	110	NA
GSE84844	SjS	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	60	(Tasaki et al., 2017)
GSE87095	RA	Methylation profiling by array	Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)	122	(Julià et al., 2017)
GSE89408	RA	Expression profiling by high throughput sequencing	Illumina HiSeq 2000 (Homo sapiens)	218	(Guo et al., 2017)
GSE90081	RA	Expression profiling by high throughput sequencing	Illumina HiSeq 2000 (Homo sapiens)	24	(Shchetynsky et al., 2017)
GSE93683	SjS	Expression profiling by array	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	48	(Tasaki et al., 2017)
GSE95065	SSc	Expression profiling by array	[HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array (HGU133A2 Hs ENTREZG 19.0.0)	33	NA
GSE10325	SLE	Expression profiling by array	[HG-U133A] Affymetrix Human Genome U133A Array	67	

Supplementary Table 1 references

Absher, D.M., Li, X., Waite, L.L., Gibson, A., Roberts, K., Edberg, J., Chatham, W.W., and Kimberly, R.P. (2013). Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations. PLoS Genet. 9, e1003678.

Ayano, M., Tsukamoto, H., Kohno, K., Ueda, N., Tanaka, A., Mitoma, H., Akahoshi, M., Arinobu, Y., Niiro, H., Horiuchi, T., et al. (2015). Increased CD226 Expression on CD8+ T Cells Is Associated with Upregulated Cytokine Production and Endothelial Cell Injury in Patients with Systemic Sclerosis. J. Immunol. 195, 892–900.

Banchereau, R., Hong, S., Cantarel, B., Baldwin, N., Baisch, J., Edens, M., Cepika, A.-M., Acs, P., Turner, J., Anguiano, E., et al. (2016). Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. Cell 165, 551–565.

Bienkowska, J., Allaire, N., Thai, A., Goyal, J., Plavina, T., Nirula, A., Weaver, M., Newman, C., Petri, M., Beckman, E., et al. (2014). Lymphotoxin-LIGHT pathway regulates the interferon signature in rheumatoid arthritis. PLoS ONE 9, e112545.

Broeren, M.G.A., de Vries, M., Bennink, M.B., Arntz, O.J., Blom, A.B., Koenders, M.I., van Lent, P.L.E.M., van der Kraan, P.M., van den Berg, W.B., and van de Loo, F.A.J. (2016). Disease-Regulated Gene Therapy with Anti-Inflammatory Interleukin-10 Under the Control of the CXCL10 Promoter for the Treatment of Rheumatoid Arthritis. Hum. Gene Ther. 27, 244–254.

Chaussabel, D., Quinn, C., Shen, J., Patel, P., Glaser, C., Baldwin, N., Stichweh, D., Blankenship, D., Li, L., Munagala, I., et al. (2008). A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity 29, 150–164.

Cole, M.B., Quach, H., Quach, D., Baker, A., Taylor, K.E., Barcellos, L.F., and Criswell, L.A. (2016). Epigenetic Signatures of Salivary Gland Inflammation in Sjögren's Syndrome. Arthritis & Rheumatology (Hoboken, N.J.) 68, 2936–2944.

Fernandez, D.R., Telarico, T., Bonilla, E., Li, Q., Banerjee, S., Middleton, F.A., Phillips, P.E., Crow, M.K., Oess, S., Muller-Esterl, W., et al. (2009). Activation of mammalian target of rapamycin controls the loss of TCRzeta in lupus T cells through HRES-1/Rab4-regulated lysosomal degradation. J. Immunol. 182, 2063–2073.

Gao, P., Uzun, Y., He, B., Salamati, S.E., Coffey, J.K.M., Tsalikian, E., and Tan, K. (2019). Risk variants disrupting enhancers of TH1 and TREG cells in type 1 diabetes. Proc. Natl. Acad. Sci. U.S.A. 116, 7581–7590.

Garaud, J.-C., Schickel, J.-N., Blaison, G., Knapp, A.-M., Dembele, D., Ruer-Laventie, J., Korganow, A.-S., Martin, T., Soulas-Sprauel, P., and Pasquali, J.-L. (2011). B cell signature during inactive systemic lupus is heterogeneous: toward a biological dissection of lupus. PLoS ONE 6, e23900.

Greenwell-Wild, T., Moutsopoulos, N.M., Gliozzi, M., Kapsogeorgou, E., Rangel, Z., Munson, P.J., Moutsopoulos, H.M., and Wahl, S.M. (2011). Chitinases in the salivary glands and circulation of patients with Sjögren's syndrome: macrophage harbingers of disease severity. Arthritis Rheum. 63, 3103–3115.

Guo, Y., Walsh, A.M., Fearon, U., Smith, M.D., Wechalekar, M.D., Yin, X., Cole, S., Orr, C., McGarry, T., Canavan, M., et al. (2017). CD40L-Dependent Pathway Is Active at Various Stages of Rheumatoid Arthritis Disease Progression. J. Immunol. 198, 4490–4501.

Hong, K.-M., Kim, H.-K., Park, S.-Y., Poojan, S., Kim, M.-K., Sung, J., Tsao, B.P., Grossman, J.M., Rullo, O.J., Woo, J.M.P., et al. (2017). CD3Z hypermethylation is associated with severe clinical manifestations in systemic lupus erythematosus and reduces CD3ζ-chain expression in T cells. Rheumatology (Oxford) 56, 467–476.

Horvath, S., Nazmul-Hossain, A.N.M., Pollard, R.P.E., Kroese, F.G.M., Vissink, A., Kallenberg, C.G.M., Spijkervet, F.K.L., Bootsma, H., Michie, S.A., Gorr, S.U., et al. (2012). Systems analysis of primary Sjögren's syndrome pathogenesis in salivary glands identifies shared pathways in human and a mouse model. Arthritis Res. Ther. 14, R238.

Hu, S., Wang, J., Meijer, J., Ieong, S., Xie, Y., Yu, T., Zhou, H., Henry, S., Vissink, A., Pijpe, J., et al. (2007). Salivary proteomic and genomic biomarkers for primary Sjögren's syndrome. Arthritis Rheum. 56, 3588–3600.

Huber, R., Hummert, C., Gausmann, U., Pohlers, D., Koczan, D., Guthke, R., and Kinne, R.W. (2008). Identification of intra-group, inter-individual, and gene-specific variances in mRNA expression profiles in the rheumatoid arthritis synovial membrane. Arthritis Res. Ther. 10, R98.

Hung, T., Pratt, G.A., Sundararaman, B., Townsend, M.J., Chaivorapol, C., Bhangale, T., Graham, R.R., Ortmann, W., Criswell, L.A., Yeo, G.W., et al. (2015). The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. Science 350, 455–459.

Hutcheson, J., Scatizzi, J.C., Siddiqui, A.M., Haines, G.K., Wu, T., Li, Q.-Z., Davis, L.S., Mohan, C., and Perlman, H. (2008). Combined deficiency of proapoptotic regulators Bim and Fas results in the early onset of systemic autoimmunity. Immunity 28, 206–217.

Jeffries, M.A., Dozmorov, M., Tang, Y., Merrill, J.T., Wren, J.D., and Sawalha, A.H. (2011). Genome-wide DNA methylation patterns in CD4+ T cells from patients with systemic lupus erythematosus. Epigenetics 6, 593–601.

Julià, A., Absher, D., López-Lasanta, M., Palau, N., Pluma, A., Waite Jones, L., Glossop, J.R., Farrell, W.E., Myers, R.M., and Marsal, S. (2017). Epigenome-wide association study of rheumatoid arthritis identifies differentially methylated loci in B cells. Hum. Mol. Genet. 26, 2803–2811.

Kennedy, W.P., Maciuca, R., Wolslegel, K., Tew, W., Abbas, A.R., Chaivorapol, C., Morimoto, A., McBride, J.M., Brunetta, P., Richardson, B.C., et al. (2015). Association of the interferon signature metric with serological disease manifestations but not global activity scores in multiple cohorts of patients with SLE. Lupus Sci Med 2, e000080.

Lessard, C.J., Li, H., Adrianto, I., Ice, J.A., Rasmussen, A., Grundahl, K.M., Kelly, J.A., Dozmorov, M.G., Miceli-Richard, C., Bowman, S., et al. (2013). Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren's syndrome. Nat. Genet. 45, 1284–1292.

Li, Q.-Z., Karp, D.R., Quan, J., Branch, V.K., Zhou, J., Lian, Y., Chong, B.F., Wakeland, E.K., and Olsen, N.J. (2011). Risk factors for ANA positivity in healthy persons. Arthritis Res. Ther. 13, R38.

Linsley, P.S., Speake, C., Whalen, E., and Chaussabel, D. (2014). Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. PLoS ONE 9, e109760.

Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., et al. (2013). Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat. Biotechnol. 31, 142–147.

Mariotti, B., Servaas, N.H., Rossato, M., Tamassia, N., Cassatella, M.A., Cossu, M., Beretta, L., van der Kroef, M., Radstake, T.R.D.J., and Bazzoni, F. (2019). The Long Non-coding RNA NRIR Drives IFN-Response in Monocytes: Implication for Systemic Sclerosis. Front Immunol 10, 100.

Moreno-Moral, A., Bagnati, M., Koturan, S., Ko, J.-H., Fonseca, C., Harmston, N., Game, L., Martin, J., Ong, V., Abraham, D.J., et al. (2018). Changes in macrophage transcriptome associate with systemic sclerosis and mediate GSDMA contribution to disease risk. Ann. Rheum. Dis. 77, 596–601.

Rai, R., Chauhan, S.K., Singh, V.V., Rai, M., and Rai, G. (2016). RNA-seq Analysis Reveals Unique Transcriptome Signatures in Systemic Lupus Erythematosus Patients with Distinct Autoantibody Specificities. PLoS ONE 11, e0166312.

Rakyan, V.K., Beyan, H., Down, T.A., Hawa, M.I., Maslau, S., Aden, D., Daunay, A., Busato, F., Mein, C.A., Manfras, B., et al. (2011). Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. PLoS Genet. 7, e1002300.

Rosenberg, A., Fan, H., Chiu, Y.G., Bolce, R., Tabechian, D., Barrett, R., Moorehead, S., Baribaud, F., Liu, H., Peffer, N., et al. (2014). Divergent gene activation in peripheral blood and tissues of patients with rheumatoid arthritis, psoriatic arthritis and psoriasis following infliximab therapy. PLoS ONE 9, e110657.

Shchetynsky, K., Diaz-Gallo, L.-M., Folkersen, L., Hensvold, A.H., Catrina, A.I., Berg, L., Klareskog, L., and Padyukov, L. (2017). Discovery of new candidate genes for rheumatoid arthritis through integration of genetic association data with expression pathway analysis. Arthritis Res. Ther. 19, 19.

Smiljanovic, B., Grün, J.R., Biesen, R., Schulte-Wrede, U., Baumgrass, R., Stuhlmüller, B., Maslinski, W., Hiepe, F., Burmester, G.-R., Radbruch, A., et al. (2012). The multifaceted balance of TNF-α and type I/II interferon responses in SLE and RA: how monocytes manage the impact of cytokines. J. Mol. Med. 90, 1295–1309.

Tasaki, S., Suzuki, K., Nishikawa, A., Kassai, Y., Takiguchi, M., Kurisu, R., Okuzono, Y., Miyazaki, T., Takeshita, M., Yoshimoto, K., et al. (2017). Multiomic disease signatures converge to cytotoxic CD8 T cells in primary Sjögren's syndrome. Ann. Rheum. Dis. 76, 1458–1466.

Tsoi, L.C., Hile, G.A., Berthier, C.C., Sarkar, M.K., Reed, T.J., Liu, J., Uppala, R., Patrick, M., Raja, K., Xing, X., et al. (2019). Hypersensitive IFN Responses in Lupus Keratinocytes Reveal Key Mechanistic Determinants in Cutaneous Lupus. J. Immunol. 202, 2121–2130.

Ulff-Møller, C.J., Asmar, F., Liu, Y., Svendsen, A.J., Busato, F., Grønbaek, K., Tost, J., and Jacobsen, S. (2018). Twin DNA Methylation Profiling Reveals Flare-Dependent Interferon Signature and B Cell Promoter Hypermethylation in Systemic Lupus Erythematosus. Arthritis & Rheumatology (Hoboken, N.J.) 70, 878–890.

Vecchio, F., Lo Buono, N., Stabilini, A., Nigi, L., Dufort, M.J., Geyer, S., Rancoita, P.M., Cugnata, F., Mandelli, A., Valle, A., et al. (2018). Abnormal neutrophil signature in the blood and pancreas of presymptomatic and symptomatic type 1 diabetes. JCI Insight 3.

Walter, G.J., Fleskens, V., Frederiksen, K.S., Rajasekhar, M., Menon, B., Gerwien, J.G., Evans, H.G., and Taams, L.S. (2016). Phenotypic, Functional, and Gene Expression Profiling of Peripheral CD45RA+ and CD45RO+ CD4+CD25+CD127(low) Treg Cells in Patients With Chronic Rheumatoid Arthritis. Arthritis & Rheumatology (Hoboken, N.J.) 68, 103–116.

Woetzel, D., Huber, R., Kupfer, P., Pohlers, D., Pfaff, M., Driesch, D., Häupl, T., Koczan, D., Stiehl, P., Guthke, R., et al. (2014). Identification of rheumatoid arthritis and osteoarthritis patients by transcriptome-based rule set generation. Arthritis Res. Ther. 16, R84.

Yang, M., Ye, L., Wang, B., Gao, J., Liu, R., Hong, J., Wang, W., Gu, W., and Ning, G. (2015). Decreased miR-146 expression in peripheral blood mononuclear cells is correlated with ongoing islet autoimmunity in type 1 diabetes patients 1miR-146. J Diabetes 7, 158–165.

Ye, H., Zhang, J., Wang, J., Gao, Y., Du, Y., Li, C., Deng, M., Guo, J., and Li, Z. (2015). CD4 T-cell transcriptome analysis reveals aberrant regulation of STAT3 and Wnt signaling pathways in rheumatoid arthritis: evidence from a case-control study. Arthritis Res. Ther. 17, 76.