**Inference of an integrative network for Rheumatoid Arthritis combining data-driven machine learning approaches and a state-of-the-art mechanistic disease map.**

Quentin Miagoux[1], Dereck de Mezquita[1], Vidisha Singh[1], Valerie Chaudru[1], Mohamed Elati[2], Elisabeth Petit-Teixeira[1], Anna Niarakis[1,3]

1 GenHotel, Univ. Évry, University of Paris-Saclay, Genopole, 91025, Évry, France
2 CANTHER, University of Lille, CNRS UMR 1277, Inserm U9020, 59045 Lille, France
3 Lifeware Group, Inria Saclay-île de France, Palaiseau 91120, France

**Motivation:** Rheumatoid arthritis (RA) is a multifactorial autoimmune disease that causes chronic inflammation of the joints. RA is considered a complex disease as it involves various genetic, environmental, and epigenetic factors. Systems biology approaches provide the means to study complex diseases by integrating different layers of biological information. Combining multiple data types can help compensate for missing or conflicting information and limit the possibility of false positives. In this approach, we integrate three different biological layers (gene expression, signalling cascades, mutations), obtained by bottom-up and top-down methods to build an integrative, disease-specific network. The goal behind this endeavour is to see if we can unravel mechanisms governing the regulation of key genes identified as mutation carriers in RA and derive patient-specific models to gain more insights into the disease heterogeneity.

**Results:** In this work, we combine biological data relevant to Rheumatoid Arthritis, in the form of a global, integrative network. We first make use of publicly available transcriptomic datasets (peripheral blood) relative to RA and machine learning to create an RA specific transcription factor (TF) co-regulatory network. The TF cooperativity network is subsequently enriched in signalling cascades and upstream regulators using prior knowledge encoded in a state-of-the-art, RA-specific molecular map. Lastly, a list of RA specific variants highlights key genes associated with known disease mutations.

**Availability:** Datasets used for the analysis are publicly available. All scripts used to generate results and the Shiny app will be freely accessible after peer-reviewed publication.

**Contact:** anna.niaraki@univ-evry.fr

**Key words:** network inference, integrative biology, rheumatoid arthritis, signalling and gene regulation

## Introduction

Rheumatoid Arthritis (RA) is an inflammatory, autoimmune disease that affects the joints of the body. While the exact aetiology is unknown, it involves a combination of environmental and genetic factors such as smoking and susceptibility genes, along with sex and age factors. RA affects 0.5-1% of the world population with women three times more susceptible to develop RA than men (Ngo *et al.*, 2014; Smolen *et al.*, 2018). The onset of the disease is set around the fourth to fifth decade of one's life (Deane *et al.*, 2017) and if left untreated can be debilitating for the individual.

Symptoms of RA include synovial inflammation, joint stiffness and pain, cartilage destruction and bone erosion. Serologically RA is associated with the presence of autoantibodies, such as the anti-citrullinated protein antibody (ACPA) and the rheumatoid factor (RF). In early RA, leukocytes invade the synovial joints followed by other pro-inflammatory mediators which instigate an inflammatory cascade and provoke synovitis (Smolen *et al.*, 2018). Activated

monocytes and T cells, both a source of proinflammatory cytokines such as TNF-a, can be found in peripheral blood (Batliwalla *et al.*, 2005) and many RA studies have used peripheral blood cells to identify disease-related genes (Edwards *et al.*, 2007; Micsik *et al.*, 2015; Kuuliala *et al.*, 2017; Li *et al.*, 2020).

The common therapy for the disease includes the use of Disease-modifying anti-rheumatic drugs (DMARDs). Conventional DMARDs include drugs that target the entire immune system, whereas biologic DMARDs are monoclonal antibodies (mAbs) and soluble receptors that target protein messenger molecules or cells. Patients who do not respond to conventional DMARDs usually initiate therapy with TNF inhibitors. However, approximately 30–40% of RA patients fail to respond to anti-TNF therapy and are usually obliged to go several rounds of drug combinations (Farutin *et al.*, 2019).

Due to the complex nature of the disease, systems biology and integrative approaches are needed in order to gain insight into the disease pathogenesis and progression. It is evident that focusing only on one aspect of the disease provides a limited understanding of the multifactorial nature of RA Recently, network-based approaches which rely on the integration of multi-omic data (proteomics, genomics, transcriptomics and metabolomics) have succeeded in unravelling key mechanisms in complex diseases (Eguchi *et al.*, 2018; Karimizadeh *et al.*, 2019; Haghjoo *et al.*, 2020; Sahu *et al.*, 2020).

Among computational approaches, machine learning is a promising bioinformatics field that allows the use of big data from a variety of biomedical datasets. Studies have shown that incorporating prior knowledge to data driven methodologies improves the quality and the biological relevance of the outcome (Greenfield *et al.*, 2013; Zuo *et al.*, 2017; Benedetti *et al.*, 2020). CoRegNet is an R/Bioconductor package that infers co-regulatory networks of transcription factors (TFs) and target genes by analysing transcriptomic data and estimating TFs activity profiles. Moreover, the software allows for network enrichment by integrating regulation evidence for TF binding sites, protein-protein interaction data and ChiP data from various databases to support cooperative TFs (Nicolle *et al.*, 2015).

In this work, we present a framework for the integration of signalling and transcriptional regulation cascades with genomic mutations combining data driven approaches with prior knowledge in the form of an integrative RA-specific network. To do so, we make use of publicly available transcriptomic data of white blood cells from patients suffering from RA, and the tool CoRegNet to infer a co-regulatory network. Next, we develop an integration pairing method to couple the RA co-regulatory network with a state of the art disease map for RA (Singh *et al.*, 2020), to enrich the cooperativity network with upstream, signalling regulators. Disease maps are comprehensive, knowledge-based representations of disease mechanisms including disease-related signalling, metabolic and gene regulatory processes supported by literature-based evidence (Ostaszewski *et al.*, 2019).

Lastly, we project on the integrative RA network public genomic data, highlighting key mutation carrier genes associated with the disease (Figure 1). The goal is to unravel mechanisms governing the regulation of key genes identified as mutation carriers in RA, spanning from upstream regulators like ligands and receptors, to TFs identified as master regulators and all associated target genes, placing the mutations in a hierarchical, functional biological context.

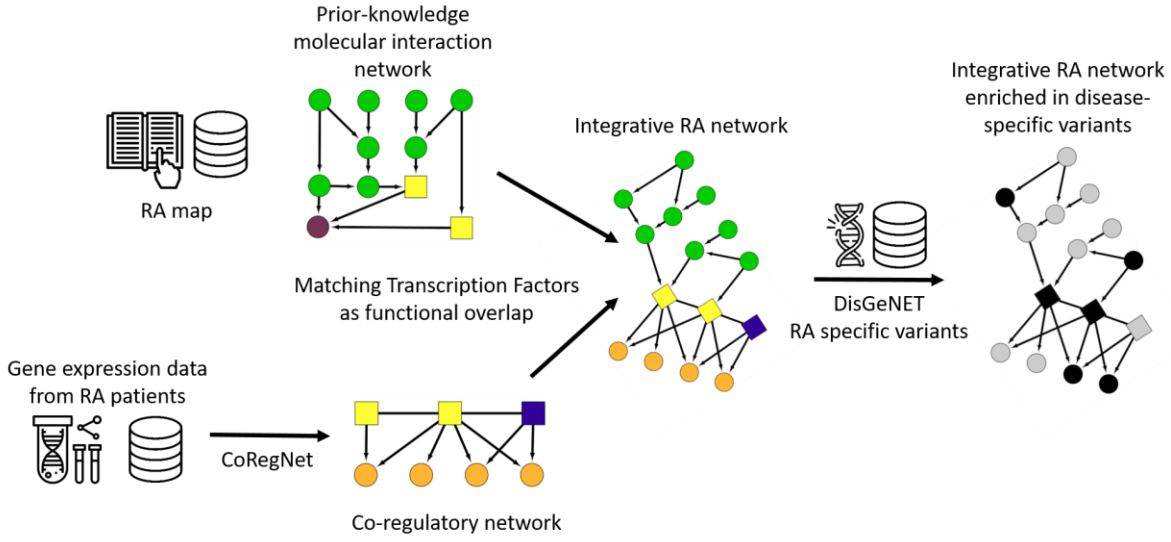


**Figure 1.** Workflow of data integration to obtain an integrative, global RA -specific network. Green round nodes denote upstream regulators of the RA map, yellow squares denote matching TFs between the RA map and the CoRegNet inferred co-regulatory network, blue squares denote cooperative nonmatching TFs. In the final, integrative network genes/proteins that are identified as mutation carriers are highlighted in black.

## Materials & Methods

### Data description and pre-processing:

We used a transcriptomic dataset from Illumina HiSeq 2500 of white blood cells from RA patients and healthy donors (GSE117769). The data set describes 120 subjects in total (51 RA and 50 control, and 19 patients with either Ankylosing spondylitis or Psoriatic arthritis). From this dataset, we extracted 96 samples (46 RA and 50 control). For the 46 RA samples, 43 were of Caucasian ancestry and 3 of unknown ancestry. Moreover, two samples consisted of duplicates of the same origin (female, control, unknown ancestry) and their expression matrix average was used for the analysis. We conducted a preliminary analysis on the expression matrix data with DESeq2 (Love *et al.*, 2014, 2), using normalisation and variance stabilising transformation on the matrix expression. Principal Component Analysis (PCA) revealed 5 outliers (4 RA and 1 control) that were removed from the dataset (Figure S1). The final dataset used for further analysis comprised 90 samples (42 RA and 48 controls). Finally, we performed a normalisation on the raw expression matrix after removal of low read counts (>10).

### Inference of the co-regulatory network (CoRegNet)

We used CoRegNet (Nicolle *et al.*, 2015) R package version 1.26.0 to infer the co-regulatory network with the normalised gene expression matrix of the pre-processing step. The CoRegNet package implements the H-LICORN algorithm, allowing the identification of cooperative regulators of genes (Chebil *et al.*, 2014). We enriched the inferred network with protein-protein and regulation evidence and further refined it with an unsupervised method using the unweighted mean. The co-regulatory network inferred with CoRegNet is composed of the significant edges between TF with a False Discovery Rate (FDR) (Benjamini and Hochberg, 1995) of 5%.

### RA map upstream protein extraction

The RA map is a state-of-the-art interactive knowledge base for the RA disease (Singh *et al.*, 2020). The RA map is organised in the form of a cell representing the flow of information from the extracellular space to the plasma membrane and then to the cytoplasm, the nucleus and the

secreted compartment or cellular phenotypes. For our analysis, as we were focused on upstream regulators of identified TFs, we used mainly the signalling part of the RA map.
More specifically, the list of TFs was uploaded as an overlay to the RA map, and the matching TFs were identified. The matching TFs were subsequently used as seeding nodes for the upstream plugin (Hoksza *et al.*, 2019), setting the mode of extraction as upstream and selecting non-blocking modifiers. The obtained file was extracted as an XMLCellDesigner file. With this method we extracted upstream signalling cascades, and also eight translation reactions for which the mRNA was directly linked with the protein in the cytoplasm or the membrane.
The RA map, and consequently the extracted network, is written in the Process Description Systems Biology Graphical Notation scheme (Le Novère *et al.*, 2009). To obtain a more simplified, Activity Flow like representation of the network, the CellDesigner XML file of the previous step was used as an input to the tool CaSQ (Aghamiri *et al.*, 2020) to create an Activity Flow like executable network. CaSQ provides SBML-qual files for performing *in silico* simulations, but in our case, we used only the SIF file that contains information about the source, interaction type, and target of the Boolean network.
The obtained SIF file was further modified to address the issue of complexes - for every complex represented as a single node in the AF network, we recreated the reactants. This way we would not miss interactions and overlaps between nodes existing inside complexes. Regarding entities represented multiple times (as genes, proteins or mRNAs), for simplification purposes we kept only one entity and merged the corresponding interactions.

**Global RA Network inference**
We used the R package igraph (Csardi and Nepusz, 2006) to convert the CoRegNet object (co-regulatory network) and the RA map SIF file into separate graphs. Then, we merged both networks using igraph, and imported the network into Cytoscape using the RCy3 R/ Bioconductor package (Gustavsen *et al.*, 2019) forming the global RA-specific network.

**List of variants**
DisGeNET (Piñero *et al.*, 2020) contains the largest publicly available collection of genes and variants associated with human diseases. From this database, we extracted 2387 variants associated with RA. Then, we filtered out variants that had a variant disease association (VDA) and evidence index (EI) score lower than 0.7. VDA score is computed using the number of curated and non-curated publications supporting the variant, while EI score is computed using contradictory results in publications supporting the variant. The use of a 0.7 threshold gives us at least one curated publication supporting the variant, as well as a good association between the variant and the disease resulting in a total of 1635 variants. Within these 1635 variants, we identified 731 associated genes that were subsequently used as an overlay for the global RA-specific network.

**Differential expression analysis (DEA)**
The 90 normalised samples of the pre-processing step were used in this step. We conducted two sets of DEA based on the biological sex and the age of the patients. In the first case, we analysed female (29 RA and 33 controls) and male (13 RA and 15 controls) patients independently using DESeq2. In the second case, we repeated with the group age >=51 (29 RA and 30 controls) and the group <=50 (13 RA and 18 controls). In both cases, we considered a gene differentially expressed if the corrected (FDR) p-value $\leq 0.05$.

**Shiny app**
The co-regulatory network inferred with CoRegNet, the RA map Activity Flow extracted network and the merged global RA network, along with their overlays, were integrated to a web-

based Shiny application using R (Chang *et al.*, 2020). The web application uses a Cytoscape viewer based on the R package cyjShiny (Shah and Shannon, 2020) and is freely available (https://quentin-miagoux.shinyapps.io/global_ra_network/).

## Results

### Inference of the co-regulatory network

The tool CoRegNet was used on normalised counts data to infer a co-regulatory network shown in Figure 2A. This network includes a total of 19 TF, 14 co-regulatory interactions and a total of 373 regulated target genes. Table 1 summarises the top five TFs with the highest number of regulatory and co-regulatory interactions. Literature search for the nineteen TF identified from CoRegNet as the master regulators in the dataset, showed their potential implication to RA Supplementary Table 1 summarises key roles of the TFs and the corresponding literature reference.

**Table 1.** Top 5 of the CoRegNet identified TFs with the highest number of regulatory interactions (TF - target gene) and co-regulatory interactions (TF-TF).

| **Top 5 TFs** | **No of regulatory interactions** |
|---|---|
| FOS | 288 |
| JUN | 211 |
| EEF1A1 | 155 |
| MNDA | 136 |
| TNFAIP3 | 125 |
| **Top 5 TFs** | **No of co-regulatory interaction(s)** |
| JUND | 5 (EEF1A1, FOS, JUN, PTMA, TNFAIP3) |
| EEF1A1 | 3 (ETS1, JUND, PTMA) |
| IRF1 | 2 (DAZAP2, FOSB) |
| MNDA | 2 (HCLS1, TRIM22) |
| PTMA | 2 (EEF1A1, JUND) |

### RA map upstream regulators of the TFs identified from CoRegNet

Six out of the 19 TFs namely ETS1, FOS, JUN, JUND, NFKBIA and TNFAIP3 from the co-regulatory network were present in the RA map and were used as seeds to extract their upstream

regulators. The extracted network comprising the RA map upstream regulators of the matching TFs includes 242 nodes and is shown in Figure 2B.

**The global, integrative RA network**

The global, integrative RA network is the result of the merging of the RA map upstream regulators and the CoRegNet object, using as interface the matching TFs. It comprises 614 nodes and 1736 interactions (848 inhibitions, 874 activations, and 14 co-regulatory interactions shared among TFs) including genes, proteins, complexes, and simple molecules shown in Figure 2C. In this network, six TFs were common between the CoRegNet network and the RA map (seeding TFs). In addition, 16 target genes identified with CoRegNet overlapped with the RA map upstream regulators.

**Integration of mutation-associated disease-specific genes as a third layer of information.**

The use of the DisGeNET variant list revealed 61 genes/proteins overlapping with the global network, as shown in Figure 3A. Among them, the TF TNFAIP3, present in both CoRegNet and RA map networks. In the RA map, TNFAIP3 is linked to TNFSF11 and IL17 as it inhibits TRAF6, an adaptor protein for these molecules. In total, TNFAIP3 participates in 128 interactions. It regulates 125 target genes from the CoRegNet network, one protein and one complex from the RA map network (TRAF6 and NFKB/NFKN respectively) and is involved in the coregulation of the TF JUND. Among TNFAIP3 targets, thirteen were also identified as mutation-carriers, namely HLA-B, TLR4, HLA-DRA, RPL37A, HLA-A, HLA-DPB1, LST1, FCGR2A, CDC42ZP3, TAGAP, IL1RN, CLEC4E and TRAF6.

**The integrative, global RA network as a template for the analysis of sex and age dependent DEGs.**

We used the global RA network as a template to project the results of DEA based on the biological sex of the RA patients in the cohort analysed. The analysis revealed 176 DEGs for male patients, 747 for female patients and 59 common for both sexes. As seen in Figure 3B, a total of 71 DEGs mapped onto the global RA network. 60 DEGs came from female patients, six from male patients and five were common for both sexes. From the five common, two of them were identified as mutation carriers, namely TLR4 and RNF213.

Regarding the analysis based on the age of the patients, we retrieved 694 for the group over 51, 280 for the group under 50 and 84 DEGs common in the two groups. As seen in Figure 3C, a total of 60 DEGs mapped onto the global RA network. 30 DEGs came from the group over 50, 17 DEGs for the group under 50 and 13 were common for both groups. From the 13 common, three were identified as mutation carriers, namely TLR4, TLR5 and IL4R.

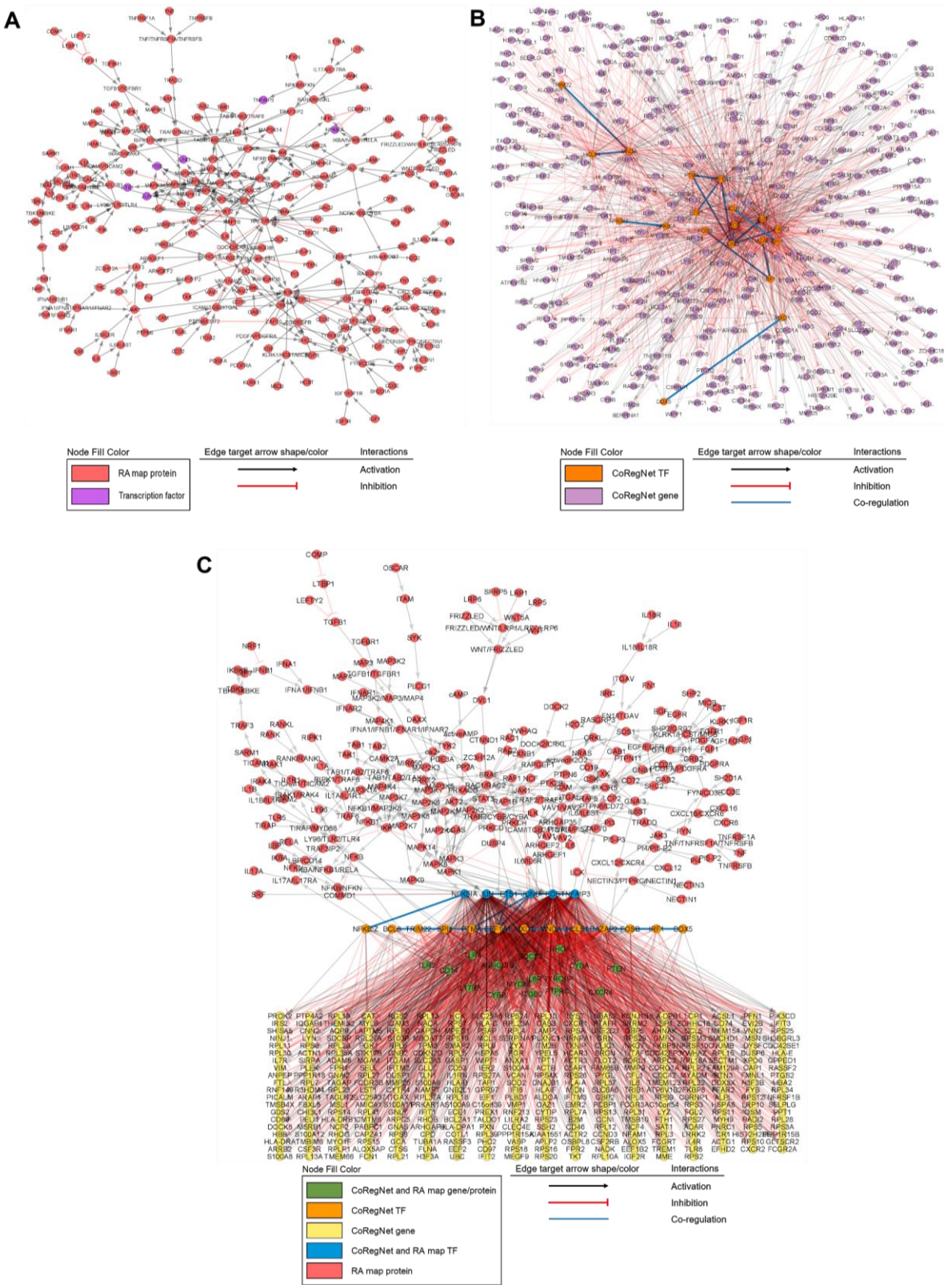


**Figure 2 A.** Co-regulatory network inferred using the tool CoRegNet and the matrix of normalised counts from the transcriptomic dataset (GSE117769). The dataset used included data from 90 patients (48 Controls and 42 RA patients). TFs from CoRegNet are coloured in orange while regulated target genes are coloured in purple **B.** The RA map upstream regulators of the matching TFs between the RA map and the CoRegNet co-regulatory network. The matching TFs are coloured in purple while the upstream regulators are coloured in red **C.** Integrative, global network for RA RA map upstream regulators are

bound to TFs identified in the CoRegNet co-regulatory network. Matching TFs are coloured in blue, matching genes/proteins are coloured in green, upstream regulators from RA map are coloured in red, CoRegNet TFs are coloured in orange and CoRegNet genes are coloured in yellow.

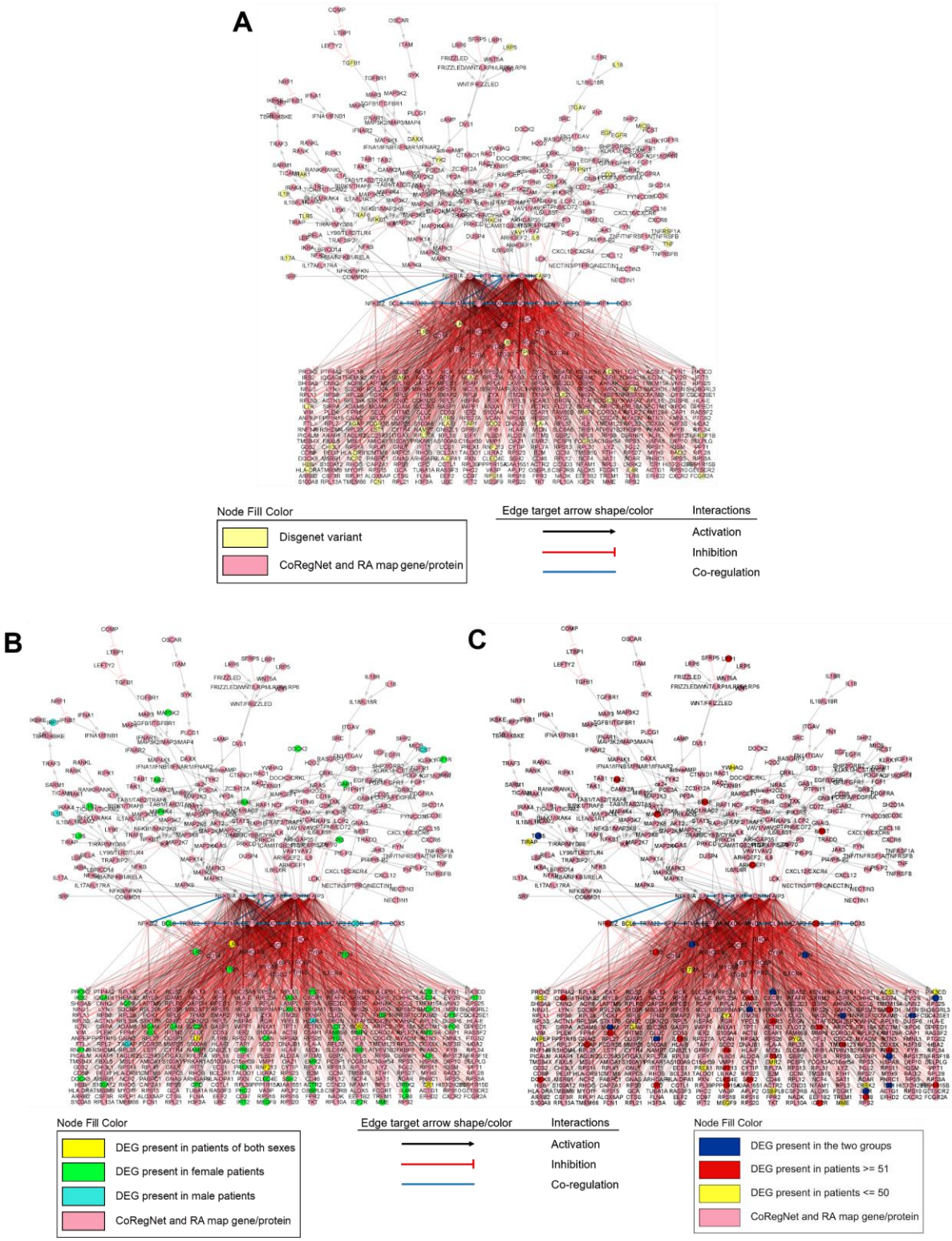


**Figure 3 A.** Global network for RA with associated RA DisGeNET variant used an overlay. Overlapping RA DisGeNET variants are shown in yellow while non-overlapping genes/proteins are shown in pink. **B**. Differentially expressed genes used as an overlay from two separate DEA based on the patients' biological sex. DEGs from female patients are represented with green nodes, DEGs from male patients with cyan

nodes, DEGs from both sexes with yellow nodes, and no overlaps with pink nodes. **C.** Differentially expressed genes used as an overlay from two separate DEA based on the patients' age. DEGs from patients with age equal and over 51years are represented with red nodes, patients with age equal, and below 50 years with yellow nodes, DEGs from both age groups with blue nodes, and no overlaps with pink nodes.

## Discussion

In this work, we present a framework to combine prior-knowledge based networks with data-driven machine learning approaches. Prior knowledge networks based on human curation remain the best solution regarding signalling cascades, while reverse engineering algorithms have shown promising performance for regulatory and co-regulatory network inference. Integrating multi-omics data in a straightforward and biological relevant way is a key step in understanding pathogenetic mechanisms of multifactorial diseases, where one level of information does not suffice to explain the complex phenotypic traits.

In our approach, we make efforts to combine bottom-up and top-down systems biology approaches to link gene coregulation by TFs, with mechanistic signalling cascades that provide insights about upstream regulation. This approach creates a sense of connectivity between signal perturbation, gene regulation and gene expression. We go one step further using this integrative, global RA network to map genetic data. Placing genes with associated mutations linked to RA into a network blueprint helps to derivate mechanistic hypothesis regarding these genes' regulation and the potential impact on the downstream expression of phenotypic traits.

The nineteen TFs identified in the co-regulatory network as master regulators have been implicated in RA and autoimmunity as indicated by the literature search, we performed to assess the biological relevance of the results. Six TFs of the co-regulatory network were also present in the RA map, a state-of-the-art mechanistic network for the disease. These six TFs, namely JUN, JUND, FOS, NFKBIA, ETS1 and TNFAIP3 have been reported in numerous studies as RA relevant and are also present in the RA map. Using them as a functional overlap between the coregulation and the signalling events, we were able to obtain a network comprising upstream cascades, active TFs and target genes. The genomic data overlay on the integrative, global RA network highlighted TNFAIP3 as a disease-relevant mutation carrier TF Our analysis showed that 127 target genes/ proteins are potentially regulated by TNFAIP3. Thirteen of these genes were also identified as disease-relevant mutation carriers.

In RA, genome-wide association studies (GWAS) have helped identify pathways that could participate in disease pathogenesis. These pathways include the CD40 signalling pathway (RA related SNPs found in CD40, TRAF1, TRAF6, TNFAIP3, and NF-κB genes) and also the T cell receptor (TCR) signalling pathway including among others the PTPN22, RasGRP, PKC-θ, TNFAIP3, and TRAF6 genes (Viatte *et al.*, 2013). Our approach helps identify the intertwined pathways that link SNPs, genes and signalling cascades as most of the indicated proteins/genes are present in the integrated network.

The integrated, global RA network was also used to visualise sex and age-related patient data. In Figure 3B, we can see the DEGs coming from male patients highlighted in yellow while data coming from female patients are highlighted in green. Out of 71 DEGs that mapped into the global RA network, 60 come from female patients, six from male patients and five are common in both sexes. From the five common, two of them are also mutation carriers (TLR4 and RNF213). While the female group is almost twice as big as the male group, the number of DEGs was sixfold higher for the female group. This is consistent with previous studies showing significant differences in transcriptome data related to the biological sex (Jansen *et al.*, 2014; Karimizadeh *et al.*, 2019) indicating that sex-bias in gene expression may underlie sex differences in the prevalence of common diseases, including RA where the women to men ratio is 3 to 1. Moreover, sexual dimorphism of the mammalian immune system is well documented, with women more susceptible

to autoimmune diseases, and men more likely to develop severe infectious diseases (Gal-Oz *et al.*, 2019). The highest number of DEGs in women could highlight the activation of different autoimmune pathways compared to men.

Regarding the age-related analysis, the age group over 51 was twice as big as the under 50, and it also showed an increased expression of DEGs (2,5 times higher than the under 50 group). RA is age-dependent, and most RA patients are postmenopausal women. The incidence of the disease is higher after menopause and continues to rise at least up to the seventh decade of life (Goronzy *et al.*, 2010; Pikwer *et al.*, 2012). As seen in Figure 3C, a total of 60 DEGs mapped onto the global RA network. 30 DEGs came from the group over 51, 17 DEGs for the group under 50 and 13 were common for both groups. From the 13 common, three were identified as mutation carriers, namely TLR4, TLR5 and IL4R.

TLR4 was identified as a DEG in both sexes and age groups and has also been identified as a disease associated mutation carrier. Despite numerous mechanistic studies suggesting TLR4 as a possible pharmaceutical target, the results of phase 2 clinical trial using NI-0101, a humanised monoclonal antibody blocking TLR4, demonstrated that blocking the TLR4 pathway alone did not improve the disease outcome. Since other TLRs like TLR2, TLR5, and TLR7 are also implicated in RA pathology, redundancy in TLR signalling could explain the poor outcome of TLR4 blockade (Monnet *et al.*, 2020). Indeed, looking at the complexity of the integrated RA network, where TLR2,4 and 5 are present, it becomes rapidly evident that the blocking or downregulation of disease-associated traits like inflammation would more likely stem from a combined perturbation effect rather than a single-point intervention.

In this study, we show how an integrative signalling and coregulation network can be used to derive mechanistic insights related to signalling and transcriptional activity and provide a template for placing reported genomic alterations of the disease in their functional context.

This type of integrated networks can help towards personalised models, as placing the mutation-associated genes of different patients or patients' cohorts, such as responders and non-responders, within the integrated network could reveal potential upstream and downstream impacted pathways that might explain or predict resistance to a specific therapy.

**Acknowledgements**

We would like to thank Dr Smahane Chalabi, GenHotel, UEVE for her comments on the dataset's statistical treatment.

**Conflict of interest**

The authors declare no conflict of interest

**Funding**

This work is funded by the doctorate program of the University of Paris Saclay, France

**Authors' contribution**

Q. M and DM analysed omic data and generated the CoRegNet object. QM and VS created the RA map upstream signalling network and performed literature search. QM created the shiny application and performed the coupling of the networks. QM, DM and VS created the figures. VC advised on the statistical treatment. ME supervised the network inference. E.P-T advised on RA genomic data and supervised the work. AN designed the study, guided analysis and supervised the work. AN, QM and VS prepared the original draft. All authors reviewed, edited and agreed to the published version of the manuscript.

## References


Aghamiri, SS *et al.* (2020) Automated inference of Boolean models from molecular interaction maps using CaSQ. *Bioinformatics*, **36**, 4473–4482.

Batliwalla,F.M. *et al.* (2005) Peripheral blood gene expression profiling in rheumatoid arthritis. *Genes Immun*, **6**, 388–397.

Benedetti,E. *et al.* (2020) A strategy to incorporate prior knowledge into correlation network cutoff selection. *Nature Communications*, **11**, 5153.

Benjamini,Y. and Hochberg,Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Chang,W. *et al.* (2020) shiny: Web Application Framework for R.

Chebil,I. *et al.* (2014) Hybrid method inference for the construction of cooperative regulatory network in human. *IEEE Trans Nanobioscience*, **13**, 97–103.

Csardi,G. and Nepusz,T. (2006) The igraph software package for complex network research. *InterJournal*, **Complex Systems**, 1695.

Deane,K.D. *et al.* (2017) Genetic and environmental risk factors for rheumatoid arthritis. *Best Pract Res Clin Rheumatol*, **31**, 3–18.

Edwards,C.J. *et al.* (2007) Molecular Profile of Peripheral Blood Mononuclear Cells from Patients with Rheumatoid Arthritis. *Mol Med*, **13**, 40–58.

Eguchi,R. *et al.* (2018) An integrative network-based approach to identify novel disease genes and pathways: a case study in the context of inflammatory bowel disease. *BMC Bioinformatics*, **19**, 264.

Farutin,V. *et al.* (2019) Molecular profiling of rheumatoid arthritis patients reveals an association between innate and adaptive cell populations and response to anti-tumor necrosis factor. *Arthritis Research & Therapy*, **21**, 216.

Gal-Oz,S.T. *et al.* (2019) ImmGen report: sexual dimorphism in the immune system transcriptome. *Nat Commun*, **10**, 4295.

Goronzy,J.J. *et al.* (2010) Immune aging and rheumatoid arthritis. *Rheum Dis Clin North Am*, **36**, 297–310.

Greenfield,A. *et al.* (2013) Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. *Bioinformatics*, **29**, 1060–1067.

Gustavsen *et al.* (2019) RCy3: Network Biology using Cytoscape from within R. *F1000Research*.

Haghjoo,N. *et al.* (2020) Introducing a panel for early detection of lung adenocarcinoma by using data integration of genomics, epigenomics, transcriptomics and proteomics. *Experimental and Molecular Pathology*, **112**, 104360.

Hoksza,D. *et al.* (2019) MINERVA API and plugins: opening molecular network analysis and visualisation to the community. *Bioinformatics*, **35**, 4496–4498.

Jansen,R. *et al.* (2014) Sex differences in the human peripheral blood transcriptome. *BMC Genomics*, **15**, 33.

Karimizadeh,E. *et al.* (2019) Analysis of gene expression profiles and protein-protein interaction networks in multiple tissues of systemic sclerosis. *BMC Med Genomics*, **12**, 199.

Kuuliala,K. *et al.* (2017) Baseline JAK phosphorylation profile of peripheral blood leukocytes, studied by whole blood phosphospecific flow cytometry, is associated with 1-year treatment response in early rheumatoid arthritis. *Arthritis Res Ther*, **19**.

Le Novère,N. *et al.* (2009) The Systems Biology Graphical Notation. *Nat Biotechnol*, **27**, 735–741.

Li,X. *et al.* (2020) Effect of IL-34 on T helper 17 cell proliferation and IL-17 secretion by peripheral blood mononuclear cells from rheumatoid arthritis patients. *Sci Rep*, **10**.

Love,M.I. *et al.* (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**.

Micsik,T. *et al.* (2015) MDR-1 and MRP-1 activity in peripheral blood leukocytes of rheumatoid arthritis patients. *Diagn Pathol*, **10**, 216.

Monnet,E. *et al.* (2020) Efficacy and safety of NI-0101, an anti-toll-like receptor 4 monoclonal antibody, in patients with rheumatoid arthritis after inadequate response to methotrexate: a

phase II study. *Annals of the Rheumatic Diseases*, **79**, 316–323.
Ngo,S.T. *et al.* (2014) Gender differences in autoimmune disease. *Frontiers in Neuroendocrinology*, **35**, 347–369.
Nicolle,R. *et al.* (2015) CoRegNet: reconstruction and integrated analysis of co-regulatory networks. *Bioinformatics*, **31**, 3066–3068.
Ostaszewski,M. *et al.* (2019) Community-driven roadmap for integrated disease maps. *Brief Bioinform*, **20**, 659–670.
Pikwer,M. *et al.* (2012) Early menopause and severity of rheumatoid arthritis in women older than 45 years. *Arthritis Research & Therapy*, **14**, R190.
Piñero,J. *et al.* (2020) The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res*, **48**, D845–D855.
Sahu,A. *et al.* (2020) Integrative network analysis identifies differential regulation of neuroimmune system in Schizophrenia and Bipolar disorder. *Brain, Behavior, & Immunity - Health*, **2**, 100023.
Shah,O. and Shannon,P. (2020) cyjShiny: cyjShiny.
Singh,V. *et al.* (2020) RA-map: building a state-of-the-art interactive knowledge base for rheumatoid arthritis. *Database (Oxford)*, **2020**.
Smolen,J.S. *et al.* (2018) Rheumatoid arthritis. *Nat Rev Dis Primers*, **4**, 18001.
Viatte,S. *et al.* (2013) Genetics and epigenetics of rheumatoid arthritis. *Nat Rev Rheumatol*, **9**, 141–153.
Zuo,Y. *et al.* (2017) Incorporating prior biological knowledge for network-based differential gene expression analysis using differentially weighted graphical LASSO. *BMC Bioinformatics*, **18**, 99.